

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 July 2001 (19.07.2001)

PCT

(10) International Publication Number  
**WO 01/51628 A2**

- (51) International Patent Classification<sup>7</sup>: **C12N 15/09**, C07K 14/435, G01N 33/574 (74) Agents: **SMITH, DeAnn, F. et al.**; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).
- (21) International Application Number: PCT/US01/00798 (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (22) International Filing Date: 10 January 2001 (10.01.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- |            |                              |    |
|------------|------------------------------|----|
| 60/176,077 | 14 January 2000 (14.01.2000) | US |
| 60/189,167 | 14 March 2000 (14.03.2000)   | US |
| 60/192,099 | 24 March 2000 (24.03.2000)   | US |
| 60/193,480 | 29 March 2000 (29.03.2000)   | US |
| 60/205,230 | 15 May 2000 (15.05.2000)     | US |
| 60/211,315 | 9 June 2000 (09.06.2000)     | US |
| 60/220,534 | 25 July 2000 (25.07.2000)    | US |
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report
- (71) Applicant: **MILLENNIUM PREDICTIVE MEDICINE, INC.** [US/US]; One Kendall Square Bldg. 700, Cambridge, MA 02139 (US).
- (72) Inventors: **LILLIE, James**; -. **XU, Yongyao**; -. **WANG, Youzhen**; -. **STEINMANN, Kathleen**; -.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/51628 A2

(54) Title: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER

(57) Abstract: The invention relates to compositions, kits, and methods for detecting, characterizing, preventing, and treating human breast cancers. A variety of novel markers are provided, wherein changes in the levels of expression of one or more of the markers is correlated with the presence of breast cancer.

NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
IDENTIFICATION, ASSESSMENT, PREVENTION,  
AND THERAPY OF BREAST CANCER

5 RELATED APPLICATIONS

The present application claims priority to U.S. provisional application serial no. 60/176,077, filed January 14, 2000, U.S. provisional application serial no. 60/189,167, filed March 14, 2000, U.S. provisional application serial no. 60/192,099, filed March 24, 2000, U.S. provisional application serial no. 60/193,480, filed March 29, 2000, U.S.  
10 provisional application serial no. 60/205,230, filed May 15, 2000, U.S. provisional application serial no. 60/211,315, filed June 9, 2000, U.S. provisional application serial no. 60/220,534, filed July 25, 2000, all of which are expressly incorporated by reference.

FIELD OF THE INVENTION

15 The field of the invention is breast cancer, including diagnosis, characterization, management, and therapy of breast cancer.

BACKGROUND OF THE INVENTION

The increased number of cancer cases reported in the United States, and, indeed,  
20 around the world, is a major concern. Currently there are only a handful of treatments available for specific types of cancer, and these provide no absolute guarantee of success. In order to be most effective, these treatments require not only an early detection of the malignancy, but a reliable assessment of the severity of the malignancy.

The incidence of breast cancer, a leading cause of death in women, has been  
25 gradually increasing in the United States over the last thirty years. In 1997, it was estimated that 181,000 new cases were reported in the U.S., and that 44,000 people would die of breast cancer (Parker *et al.*, 1997, *CA Cancer J. Clin.* 47:5-27; Chu *et al.*, 1996, *J. Nat. Cancer Inst.* 88:1571-1579). While the pathogenesis of breast cancer is unclear, transformation of normal breast epithelium to a malignant phenotype may be  
30 the result of genetic factors, especially in women under 30 (Miki *et al.*, 1994, *Science*, 266:66-71). The discovery and characterization of *BRCA1* and *BRCA2* has recently expanded our knowledge of genetic factors which can contribute to familial breast

- 2 -

cancer. Germ-line mutations within these two loci are associated with a 50 to 85% lifetime risk of breast and/or ovarian cancer (Casey, 1997, *Curr. Opin. Oncol.* 9:88-93; Marcus *et al*, 1996, *Cancer* 77:697-709). However, it is likely that other, non-genetic factors also have a significant effect on the etiology of the disease. Regardless of its origin, breast cancer morbidity and mortality increases significantly if it is not detected early in its progression. Thus, considerable effort has focused on the early detection of cellular transformation and tumor formation in breast tissue.

Currently, the principal manner of identifying breast cancer is through detection of the presence of dense tumorous tissue. This may be accomplished to varying degrees of effectiveness by direct examination of the outside of the breast, or through mammography or other X-ray imaging methods (Jatoi, 1999, *Am. J. Surg.* 177:518-524). The latter approach is not without considerable cost, however. Every time a mammogram is taken, the patient incurs a small risk of having a breast tumor induced by the ionizing properties of the radiation used during the test. In addition, the process is expensive and the subjective interpretations of a technician can lead to imprecision, *e.g.*, one study showed major clinical disagreements for about one-third of a set of mammograms that were interpreted individually by a surveyed group of radiologists. Moreover, many women find that undergoing a mammogram is a painful experience. Accordingly, the National Cancer Institute has not recommended mammograms for women under fifty years of age, since this group is not as likely to develop breast cancers as are older women. It is compelling to note, however, that while only about 22% of breast cancers occur in women under fifty, data suggests that breast cancer is more aggressive in pre-menopausal women.

It would therefore be beneficial to provide specific methods and reagents for the diagnosis, staging, prognosis, monitoring, and treatment of diseases associated with breast cancer, or to indicate a predisposition to such for preventative measures.

#### SUMMARY OF THE INVENTION

The invention relates to novel genes associated with breast cancer as well as methods of assessing whether a patient is afflicted with breast cancer. The methods of the present invention comprise the step of comparing the level of expression of a marker in a patient sample, wherein the marker is listed in Tables 1-6 and the normal level of

expression of the marker in a control, *e.g.*, a sample from a patient without breast cancer. A significant difference between the level of expression of the marker in the patient sample and the normal level is an indication that the patient is afflicted with breast cancer. Preferably, a protein corresponding to the marker is a secreted protein or  
5 is predicted to correspond to a secreted protein. Alternatively, the marker can correspond to a protein having an extracellular portion, to one which is normally expressed in breast tissue at a detectable level, or both.

In one method, the marker(s) are preferably selected such that the positive predictive value of the method is at least about 10%. Also preferred are embodiments of  
10 the method wherein the marker is over- or under-expressed by at least two-fold in at least about 20% of stage 0 breast cancer patients, stage I breast cancer patients, stage IIA breast cancer patients, stage IIB breast cancer patients, stage IIIA breast cancer patients, stage IIIB breast cancer patients, stage IV breast cancer patients, grade I breast cancer patients, grade II breast cancer patients, grade III breast cancer patients, malignant breast  
15 cancer patients, ductal carcinoma breast cancer patients, and lobular carcinoma breast cancer patients.

In one embodiment of the methods of the present invention, the patient sample is a breast tissue-associated body fluid. Such fluids include, for example, blood fluids, lymph and cystic fluids, as well as nipple aspirates. In another embodiment, the sample  
20 comprises cells obtained from the patient. In another embodiment, the patient sample is *in vivo*.

In accordance with the methods of the present invention, the level of expression of the marker in a sample can be assessed, for example, by detecting the presence in the sample of:

- 25 • a protein or a fragment of the protein corresponding to the marker (*e.g.* using a reagent, such as an antibody, an antibody derivative, or an antibody fragment, which binds specifically with the protein or a fragment of the protein)
- a metabolite which is produced directly (*i.e.*, catalyzed) or indirectly by a  
30 protein corresponding to the marker
- a transcribed polynucleotide (*e.g.* an mRNA or a cDNA), or fragment thereof, having at least a portion with which the marker is substantially

- 4 -

homologous (*e.g.* by contacting a mixture of transcribed polynucleotides obtained from the sample with a substrate having one or more of the markers listed in Tables 1-6 fixed thereto at selected positions)

- a transcribed polynucleotide or fragment thereof, wherein the polynucleotide anneals with the marker under stringent hybridization conditions.

The methods of the present invention are particularly useful for patients with an identified breast mass or symptoms associated with breast cancer. The methods of the present invention can also be of particular use with patients having an enhanced risk of developing breast cancer (*e.g.*, patients having a familial history of breast cancer, patients identified as having a mutant oncogene, and patients at least about 50 years of age). The methods of the present invention may further be of particular use in monitoring the efficacy of treatment of a breast cancer patient (*e.g.* the efficacy of chemotherapy).

The methods of the present invention may be performed using a plurality (*e.g.* 2, 3, 5, or 10 or more) of markers. According to a method involving a plurality of markers, the level of expression in the sample of each of a plurality of markers independently selected from the markers listed in Tables 1-6 is compared with the normal level of expression of each of the plurality of markers in samples of the same type obtained from control humans not afflicted with breast cancer. A significantly enhanced level of expression of one or more of the markers listed in Tables 1-6 in the sample, relative to the corresponding normal levels, is an indication that the patient is afflicted with breast cancer. The markers of Tables 1-6 may also be used in combination with known breast cancer markers in the methods of the present invention.

In a preferred method of assessing whether a patient is afflicted with breast cancer (*e.g.*, new detection ("screening"), detection of recurrence, reflex testing), the method comprises comparing:

- a) the level of expression of a marker in a patient sample, wherein at least one marker is selected from the markers of Tables 1-6, and
- b) the normal level of expression of the marker in a control non-breast cancer sample.

- 5 -

A significant difference between the level of expression of the marker in the patient sample and the normal level is an indication that the patient is afflicted with breast cancer.

The methods of the present invention further include a method of assessing the efficacy of a test compound for inhibiting breast cancer in a patient. This method comprises comparing:

- a) expression of a marker in a first sample obtained from the patient and maintained in the presence of the test compound, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6, and
- b) expression of the marker in a second sample obtained from the patient and maintained in the absence of the test compound.

A significantly lower level of expression of the marker in the first sample, relative to the second sample, is an indication that the test compound is efficacious for inhibiting breast cancer in the patient. For example, the first and second samples can be portions of a single sample obtained from the patient or portions of pooled samples obtained from the patient.

The invention further relates to a method of assessing the efficacy of a therapy for inhibiting breast cancer in a patient. This method comprises comparing:

- a) expression of a marker in a first sample obtained from the patient prior to providing at least a portion of the therapy to the patient, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6, and
- b) expression of the marker in a second sample obtained from the patient following provision of the portion of the therapy.

A significantly lower level of expression of the marker in the second sample, relative to the first sample, is an indication that the therapy is efficacious for inhibiting breast cancer in the patient.

It will be appreciated that in these methods the "therapy" may be any therapy for treating breast cancer including, but not limited to, chemotherapy, radiation therapy and surgical removal of tissue, *e.g.*, a breast tumor. Thus, the methods of the invention may

be used to evaluate a patient before, during and after therapy, for example, to evaluate the reduction in tumor burden.

The present invention therefore further comprises a method for monitoring the progression of breast cancer in a patient, the method comprising:

- 5 a) detecting in a patient sample at a first time point, the expression of a marker, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6;
- b) repeating step a) at a subsequent time point; and
- c) comparing the level of expression detected in steps a) and b), and therefrom
- 10 monitoring the progression of breast cancer in the patient.

The invention also includes a method of selecting a composition for inhibiting breast cancer in a patient. This method comprises the steps of:

- a) obtaining a sample comprising cancer cells from the patient;
- b) separately maintaining aliquots of the sample in the presence of a
- 15 plurality of test compositions;
- c) comparing expression of a marker listed in Tables 1-6 in each of the aliquots; and
- d) selecting one of the test compositions which induces a lower level of expression of the marker in the aliquot containing that test composition,
- 20 relative to other test compositions.

In addition, the invention includes a method of inhibiting breast cancer in a patient. This method comprises the steps of:

- a) obtaining a sample comprising cancer cells from the patient;
- b) separately maintaining aliquots of the sample in the presence of a
- 25 plurality of test compositions;
- c) comparing expression of a marker listed in Tables 1-6 in each of the aliquots; and
- d) administering to the patient at least one of the test compositions which induces a lower level of expression of the marker in the aliquot
- 30 containing that test composition, relative to other test compositions.

The invention also includes a kit for assessing whether a patient is afflicted with breast cancer. This kit comprises reagents for assessing expression of a marker listed in Tables 1-6.

5 In another aspect, the invention relates to a kit for assessing the suitability of each of a plurality of compounds for inhibiting breast cancer in a patient. The kit comprises a reagent for assessing expression of a marker listed in Tables 1-6, and may also comprise a plurality of compounds.

10 In another aspect, the invention relates to a kit for assessing the presence of breast cancer cells. This kit comprises an antibody, wherein the antibody binds specifically with a protein corresponding to a marker listed in Tables 1-6. The kit may also comprise a plurality of antibodies, wherein the plurality binds specifically with a protein corresponding to a different marker which is also listed in Tables 1-6.

The invention also includes a kit for assessing the presence of breast cancer cells, wherein the kit comprises a nucleic acid probe. The probe binds specifically with a  
15 transcribed polynucleotide corresponding to a marker listed in Tables 1-6. The kit may also comprise a plurality of probes, wherein each of the probes binds specifically with a transcribed polynucleotide corresponding to a different marker listed in Tables 1-6.

The invention further relates to a method of making an isolated hybridoma which produces an antibody useful for assessing whether a patient is afflicted with breast  
20 cancer. The method comprises isolating a protein or protein fragment corresponding to a marker listed in Tables 1-6, immunizing a mammal using the isolated protein or protein fragment, isolating splenocytes from the immunized mammal, fusing the isolated splenocytes with an immortalized cell line to form hybridomas, and screening individual hybridomas for production of an antibody which specifically binds with the protein or  
25 protein fragment to isolate the hybridoma. The invention also includes an antibody produced by this method.

The invention further includes a method of assessing the breast carcinogenic or irregular growth promoting potential of a test compound. This method comprises the steps of:

- 30 a) maintaining separate aliquots of breast cells in the presence and absence of the test compound; and  
b) comparing expression of a marker in each of the aliquots.



The marker is selected from those listed in Tables 1-6. A significantly enhanced level of expression of the marker in the aliquot maintained in the presence of (or exposed to) the test compound, relative to the aliquot maintained in the absence of the test compound, is an indication that the test compound possesses breast carcinogenic or irregular growth promoting potential.

Additionally, the invention includes a kit for assessing the breast carcinogenic potential of a test compound. The kit comprises breast cells and a reagent for assessing expression of a marker in each of the aliquots. The marker is selected from those listed in Tables 1-6.

The invention further includes a method of treating a patient afflicted with breast cancer, comprising providing to cells of the patient an antisense oligonucleotide complementary to a polynucleotide corresponding to a marker listed in Tables 1-6.

The invention includes a method of inhibiting breast cancer in a patient at risk for developing breast cancer. This method comprises inhibiting expression or overexpression of a gene corresponding to a marker listed in Tables 1-6.

It will be appreciated that the methods and kits of the present invention may also include known cancer markers including known breast cancer markers. It will further be appreciated that the methods and kits may be used to identify cancers other than breast cancer.

#### DETAILED DESCRIPTION OF THE INVENTION

The invention relates to newly discovered correlations between expression of certain markers and the cancerous state of breast cells. It has been discovered that the level of expression of individual markers and combinations of markers described herein correlates with the presence of breast cancer in a patient. Methods are provided for detecting the presence of breast cancer in a sample, the absence of breast cancer in a sample, the stage of breast cancer, and other characteristics of breast cancer that are relevant to prevention, diagnosis, characterization, and therapy of breast cancer in a patient.

### Definitions

As used herein, each of the following terms has the meaning associated with it in this section.

The articles "a" and "an" are used herein to refer to one or to more than one (*i.e.* to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

A "marker" is a naturally-occurring polymer corresponding to at least one of the novel nucleic acids listed in Tables 1-6. For example, markers include, without limitation, sense and anti-sense strands of genomic DNA (*i.e.* including any introns occurring therein), RNA generated by transcription of genomic DNA (*i.e.* prior to splicing), RNA generated by splicing of RNA transcribed from genomic DNA, and proteins generated by translation of spliced RNA (*e.g.* including proteins both before and after cleavage of normally cleaved regions such as transmembrane signal sequences). As used herein, "marker" may also include a cDNA made by reverse transcription of an RNA generated by transcription of genomic DNA (including spliced RNA).

As used herein a "polynucleotide corresponds to" another (a first) polynucleotide if it is related to the first polynucleotide by any of the following relationships: 1) The second polynucleotide comprises the first polynucleotide and the second polynucleotide encodes a gene product. 2) The second polynucleotide is 5' or 3' to the first polynucleotide in cDNA, RNA, genomic DNA, or fragment of any of these polynucleotides. For example, a second polynucleotide may be fragment of a gene that includes the first and second polynucleotides. The first and second polynucleotides are related in that they are components of the gene coding for a gene product, such as a protein or antibody. However, it is not necessary that the second polynucleotide comprises or overlaps with the first polynucleotide to be encompassed within the definition of "corresponding to" as used herein. For example, the first polynucleotide may be a fragment of a 3' untranslated region of the second polynucleotide. The first and second polynucleotide may be fragments of a gene coding for a gene product. The second polynucleotide may be an exon of the gene while the first polynucleotide may be an intron of the gene. 3) The second polynucleotide is the complement of the first polynucleotide.

The term "probe" refers to any molecule which is capable of selectively binding to a specifically intended target molecule, for example a marker of the invention.

Probes can be either synthesized by one skilled in the art, or derived from appropriate biological preparations. For purposes of detection of the target molecule, probes may be specifically designed to be labeled, as described herein. Examples of molecules that can be utilized as probes include, but are not limited to, RNA, DNA, proteins, antibodies, and organic monomers.

A "breast-associated" body fluid is a fluid which, when in the body of a patient, contacts or passes through breast cells or into which cells, nucleic acids or proteins shed from breast cells are capable of passing. Exemplary breast-associated body fluids include blood fluids, lymph, cystic fluid, urine and nipple aspirates.

The "normal" level of expression of a marker is the level of expression of the marker in breast cells of a patient, *e.g.* a human, not afflicted with breast cancer.

"Over-expression" and "under-expression" of a marker refer to expression of the marker of a patient at a greater or lesser level, respectively, than normal level of expression of the marker (*e.g.* at least two-fold greater or lesser level).

As used herein, the term "promoter/regulatory sequence" means a nucleic acid sequence which is required for expression of a gene product operably linked to the promoter/regulatory sequence. In some instances, this sequence may be the core promoter sequence and in other instances, this sequence may also include an enhancer sequence and other regulatory elements which are required for expression of the gene product. The promoter/regulatory sequence may, for example, be one which expresses the gene product in a tissue-specific manner.

A "constitutive" promoter is a nucleotide sequence which, when operably linked with a polynucleotide which encodes or specifies a gene product, causes the gene product to be produced in a living human cell under most or all physiological conditions of the cell.

An "inducible" promoter is a nucleotide sequence which, when operably linked with a polynucleotide which encodes or specifies a gene product, causes the gene product to be produced in a living human cell substantially only when an inducer which corresponds to the promoter is present in the cell.

A "tissue-specific" promoter is a nucleotide sequence which, when operably linked with a polynucleotide which encodes or specifies a gene product, causes the gene product to be produced in a living human cell substantially only if the cell is a cell of the tissue type corresponding to the promoter.

5 A "transcribed polynucleotide" is a polynucleotide (e.g. an RNA, a cDNA, or an analog of one of an RNA or cDNA) which is complementary to or homologous with all or a portion of a mature RNA made by transcription of a genomic DNA corresponding to a marker of the invention and normal post-transcriptional processing (e.g. splicing), if any, of the transcript.

10 "Complementary" refers to the broad concept of sequence complementarity between regions of two nucleic acid strands or between two regions of the same nucleic acid strand. It is known that an adenine residue of a first nucleic acid region is capable of forming specific hydrogen bonds ("base pairing") with a residue of a second nucleic acid region which is antiparallel to the first region if the residue is thymine or uracil.

15 Similarly, it is known that a cytosine residue of a first nucleic acid strand is capable of base pairing with a residue of a second nucleic acid strand which is antiparallel to the first strand if the residue is guanine. A first region of a nucleic acid is complementary to a second region of the same or a different nucleic acid if, when the two regions are arranged in an antiparallel fashion, at least one nucleotide residue of the first region is

20 capable of base pairing with a residue of the second region. Preferably, the first region comprises a first portion and the second region comprises a second portion, whereby, when the first and second portions are arranged in an antiparallel fashion, at least about 50%, and preferably at least about 75%, at least about 90%, or at least about 95% of the nucleotide residues of the first portion are capable of base pairing with nucleotide

25 residues in the second portion. More preferably, all nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion.

"Homologous" as used herein, refers to nucleotide sequence similarity between two regions of the same nucleic acid strand or between regions of two different nucleic acid strands. When a nucleotide residue position in both regions is occupied by the

30 same nucleotide residue, then the regions are homologous at that position. A first region is homologous to a second region if at least one nucleotide residue position of each region is occupied by the same residue. Homology between two regions is expressed in

terms of the proportion of nucleotide residue positions of the two regions that are occupied by the same nucleotide residue. By way of example, a region having the nucleotide sequence 5'-ATTGCC-3' and a region having the nucleotide sequence 5'-TATGGC-3' share 50% homology. Preferably, the first region comprises a first portion and the second region comprises a second portion, whereby, at least about 50%, and preferably at least about 75%, at least about 90%, or at least about 95% of the nucleotide residue positions of each of the portions are occupied by the same nucleotide residue. More preferably, all nucleotide residue positions of each of the portions are occupied by the same nucleotide residue.

10 A marker is "fixed" to a substrate if it is covalently or non-covalently associated with the substrate such the substrate can be rinsed with a fluid (*e.g.* standard saline citrate, pH 7.4) without a substantial fraction of the marker dissociating from the substrate.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or  
15 DNA molecule having a nucleotide sequence that occurs in nature (*e.g.* encodes a natural protein).

Expression of a marker in a patient is "significantly" higher or lower than the normal level of expression of a marker if the level of expression of the marker is greater or less, respectively, than the normal level by an amount greater than the standard error  
20 of the assay employed to assess expression, and preferably at least twice, and more preferably three, four, five or ten times that amount. Alternately, expression of the marker in the patient can be considered "significantly" higher or lower than the normal level of expression if the level of expression is at least about two, and preferably at least about three, four, or five times, higher or lower, respectively, than the normal level of  
25 expression of the marker.

Breast cancer is "inhibited" if at least one symptom of the cancer is alleviated, terminated, slowed, or prevented. As used herein, breast cancer is also "inhibited" if recurrence or metastasis of the cancer is reduced, slowed, delayed, or prevented.

A kit is any manufacture (*e.g.* a package or container) comprising at least one  
30 reagent, *e.g.* a probe, for specifically detecting a marker of the invention, the manufacture being promoted, distributed, or sold as a unit for performing the methods of the present invention.

Description

The present invention is based, in part, on identification of novel markers which are expressed at a different level in breast cancer cells than they are in normal (*i.e.* non-cancerous) breast cells. The markers of the invention correspond to nucleic acid and polypeptide molecules which can be detected in one or both of normal and cancerous breast cells. The presence, absence, or level of expression of one or more of these markers in breast cells is herein correlated with the cancerous state of the tissue. The invention thus includes compositions, kits, and methods for assessing the cancerous state of breast cells (*e.g.* cells obtained from a human, cultured human cells, archived or preserved human cells and *in vivo* cells).

The compositions, kits, and methods of the invention have the following uses, among others:

- 1) assessing whether a patient is afflicted with breast cancer;
- 2) assessing the stage of breast cancer in a human patient;
- 3) assessing the grade of breast cancer in a patient;
- 4) assessing the benign or malignant nature of breast cancer in a patient;
- 5) assessing the histological type of neoplasm (*e.g.* ductal, lobular, etc.) associated with breast cancer in a patient;
- 6) making an isolated hybridoma which produces an antibody useful for assessing whether a patient is afflicted with breast cancer;
- 7) assessing the presence of breast cancer cells;
- 8) assessing the efficacy of one or more test compounds for inhibiting breast cancer in a patient;
- 9) assessing the efficacy of a therapy for inhibiting breast cancer in a patient;
- 10) monitoring the progression of breast cancer in a patient;
- 11) selecting a composition or therapy for inhibiting breast cancer in a patient;
- 12) treating a patient afflicted with breast cancer;
- 13) inhibiting breast cancer in a patient;
- 14) assessing the carcinogenic potential of a test compound; and

- 15) inhibiting breast cancer in a patient at risk for developing breast cancer.

The invention thus includes a method of assessing whether a patient is afflicted with breast cancer. This method comprises comparing the level of expression of a marker in a patient sample and the normal level of expression of the marker in a control, *e.g.*, a non-breast cancer sample. A significant difference between the level of expression of the marker in the patient sample and the normal level is an indication that the patient is afflicted with breast cancer. The marker is selected from the group consisting of the markers listed in Tables 1-6.

10 The polynucleotides set forth in Tables 1-6 represent previously unidentified nucleotide sequences. These nucleotide sequences were identified through subtracted library experiments described herein. In Tables 1 and 3, SEQ ID NOS 316-470, 793-890, 1255-1363, 2125-2454 and 3352-3626 are preferred and SEQ ID NOS 1-315, 676-792, 1056-1254, 1645-2124 and 2942-3351 are most preferred. In Tables 2 and 4, SEQ ID NOS: 1879-1959 are preferred and SEQ ID NOS: 1-1878 are most preferred. Also provided by this invention are polynucleotides that correspond to the polynucleotides of Tables 1-6. In one embodiment, these polynucleotides are obtained by identification of a larger fragment or full-length coding sequence of these polynucleotides. Gene delivery vehicles, host cells, compositions and databases (all describe herein) containing these polynucleotides are also provided by this invention.

The invention also encompasses polynucleotides which differ from that of the polynucleotides described above, but which produce the same phenotypic effect, *e.g.* allelic variants. These altered, but phenotypically equivalent polynucleotides are referred to "equivalent nucleic acids." This invention also encompasses polynucleotides characterized by changes in non-coding regions that do not alter the polypeptide produced therefrom when compared to the polynucleotide herein. This invention further encompasses polynucleotides, which hybridize to the polynucleotides of the subject invention under conditions of moderate or high stringency. Alternatively, the polynucleotides are at least 85%, or at least 90%, or more preferably, greater or equal to 95% identical as determined by a sequence alignment program when run under default parameters.

Any marker or combination of markers listed in Tables 1-6, as well as any known markers in combination with the markers set forth in Tables 1-6, may be used in the compositions, kits, and methods of the present invention. In general, it is preferable to use markers for which the difference between the level of expression of the marker in breast cancer cells and the level of expression of the same marker in normal breast cells is as great as possible. Although this difference can be as small as the limit of detection of the method for assessing expression of the marker, it is preferred that the difference be at least greater than the standard error of the assessment method, and preferably a difference of at least 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 15-, 20-, 25-, 100-, 500-, 1000-fold or greater.

It is recognized that certain markers correspond to proteins which are secreted from breast cells (*i.e.* one or both of normal and cancerous cells) to the extracellular space surrounding the cells. These markers are preferably used in certain embodiments of the compositions, kits, and methods of the invention, owing to the fact that the protein corresponding to each of these markers can be detected in an breast-associated body fluid sample, which may be more easily collected from a human patient than a tissue biopsy sample. In addition, preferred *in vivo* techniques for detection of a protein corresponding to a marker of the invention include introducing into a subject a labeled antibody directed against the protein. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

Although not every marker corresponding to a secreted protein is indicated as such herein, it is a simple matter for the skilled artisan to determine whether any particular marker corresponds to a secreted protein. In order to make this determination, the protein corresponding to a marker is expressed in a test cell (*e.g.* a cell of a breast cell line), extracellular fluid is collected, and the presence or absence of the protein in the extracellular fluid is assessed (*e.g.* using a labeled antibody which binds specifically with the protein).

The following is an example of a method which can be used to detect secretion of a protein corresponding to a marker of the invention. About  $8 \times 10^5$  293T cells are incubated at 37°C in wells containing growth medium (Dulbecco's modified Eagle's medium {DMEM} supplemented with 10% fetal bovine serum) under a 5% (v/v) CO<sub>2</sub>,



95% air atmosphere to about 60-70% confluence. The cells are then transfected using a standard transfection mixture comprising 2 micrograms of DNA comprising an expression vector encoding the protein and 10 microliters of LipofectAMINE™ (GIBCO/BRL Catalog no. 18342-012) per well. The transfection mixture is maintained  
5 for about 5 hours, and then replaced with fresh growth medium and maintained in an air atmosphere. Each well is gently rinsed twice with DMEM which does not contain methionine or cysteine (DMEM-MC; ICN Catalog no. 16-424- 54). About 1 milliliter of DMEM-MC and about 50 microcuries of Trans-<sup>35</sup>S™ reagent (ICN Catalog no. 51006) are added to each well. The wells are maintained under the 5% CO<sub>2</sub> atmosphere  
10 described above and incubated at 37°C for a selected period. Following incubation, 150 microliters of conditioned medium is removed and centrifuged to remove floating cells and debris. The presence of the protein in the supernatant is an indication that the protein is secreted.

Examples of breast-associated body fluids include blood fluids (*e.g.* whole blood,  
15 blood serum, blood having platelets removed therefrom, etc.), lymph, ascitic fluid, cystic fluid, urine and nipple aspirates. In these embodiments, the level of expression of the marker can be assessed by assessing the amount (*e.g.* absolute amount or concentration) of the marker in a breast-associated body fluid obtained from a patient. The fluid can, of course, be subjected to a variety of well-known post-collection preparative and storage  
20 techniques (*e.g.* storage, freezing, ultrafiltration, concentration, evaporation, centrifugation, etc.) prior to assessing the amount of the marker in the fluid.

Many breast-associated body fluids (*i.e.* usually excluding urine) can have breast cells therein, particularly when the breast cells are cancerous, and, more particularly, when the breast cancer is metastasizing. Thus, the compositions, kits, and methods of  
25 the invention can be used to detect expression of markers corresponding to proteins having at least one portion which is displayed on the surface of cells which express it. It is a simple matter for the skilled artisan to determine whether the protein corresponding to any particular marker comprises a cell-surface protein. For example, immunological methods may be used to detect such proteins on whole cells, or well  
30 known computer-based sequence analysis methods (*e.g.* the SIGNALP program; Nielsen *et al.*, 1997, *Protein Engineering* 10:1-6) may be used to predict the presence of at least one extracellular domain (*i.e.* including both secreted proteins and proteins having at

least one cell-surface domain). Expression of a marker corresponding to a protein having at least one portion which is displayed on the surface of a cell which expresses it may be detected without necessarily lysing the cell (*e.g.* using a labeled antibody which binds specifically with a cell-surface domain of the protein).

- 5        Expression of a marker of the invention may be assessed by any of a wide variety of well known methods for detecting expression of a transcribed molecule or protein. Non-limiting examples of such methods include immunological methods for detection of secreted, cell-surface, cytoplasmic, or nuclear proteins, protein purification methods, protein function or activity assays, nucleic acid hybridization methods, nucleic acid
- 10    reverse transcription methods, and nucleic acid amplification methods.

- In a preferred embodiment, expression of a marker is assessed using an antibody (*e.g.* a radio-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled antibody), an antibody derivative (*e.g.* an antibody conjugated with a substrate or with the protein or ligand of a protein-ligand pair {*e.g.* biotin-streptavidin} ), or an antibody
- 15    fragment (*e.g.* a single-chain antibody, an isolated antibody hypervariable domain, etc.) which binds specifically with a protein or a fragment thereof, corresponding to the marker, such as the protein encoded by the open reading frame corresponding to the marker or such a protein which has undergone all or a portion of its normal post-translational modification.

- 20        In another preferred embodiment, expression of a marker is assessed by preparing mRNA/cDNA (*i.e.* a transcribed polynucleotide) from cells in a patient sample, and by hybridizing the mRNA/cDNA with a reference polynucleotide which is a complement of a polynucleotide comprising the marker, and fragments thereof. cDNA can, optionally, be amplified using any of a variety of polymerase chain reaction methods prior to
- 25    hybridization with the reference polynucleotide; preferably, it is not amplified.

- Expression of one or more markers can likewise be detected using quantitative PCR to assess the level of expression of the marker(s). Alternatively, any of the many known methods of detecting mutations or variants (*e.g.* single nucleotide polymorphisms, deletions, etc.) of a marker of the invention may be used to detect occurrence of a
- 30    marker in a patient.

In a related embodiment, a mixture of transcribed polynucleotides obtained from the sample is contacted with a substrate having fixed thereto a polynucleotide complementary to or homologous with at least a portion (*e.g.* at least 7, 10, 15, 20, 25, 30, 40, 50, 100, 500, or more nucleotide residues) of a marker of the invention. If polynucleotides complementary to or homologous with are differentially detectable on the substrate (*e.g.* detectable using different chromophores or fluorophores, or fixed to different selected positions), then the levels of expression of a plurality of markers can be assessed simultaneously using a single substrate (*e.g.* a "gene chip" microarray of polynucleotides fixed at selected positions). When a method of assessing marker expression is used which involves hybridization of one nucleic acid with another, it is preferred that the hybridization be performed under stringent hybridization conditions.

Because the compositions, kits, and methods of the invention rely on detection of a difference in expression levels of one or more markers of the invention, it is preferable that the level of expression of the marker is significantly greater than the minimum detection limit of the method used to assess expression in at least one of normal breast cells and cancerous breast cells.

It is understood that by routine screening of additional patient samples using one or more of the markers of the invention, it will be realized that certain of the markers are over- or under-expressed in cancers of various types, including specific breast cancers, as well as other cancers such as ovarian cancer, cervical cancer, etc. For example, it will be confirmed that some of the markers of the invention are over- or under-expressed in most (*i.e.* 50% or more) or substantially all (*i.e.* 80% or more) of breast cancer. Furthermore, it will be confirmed that certain of the markers of the invention are associated with breast cancer of various stages (*i.e.* stage 0, I, II, III, and IV breast cancers, as well as subclassifications IIA, IIB, IIIA, and IIIB, using the FIGO Stage Grouping system for primary carcinoma of the breast; (see Breast, In: *American Joint Committee on Cancer: AJCC Cancer Staging Manual*. Lippincott-Raven Publishers, 5th ed., 1997, pp. 171-180), of various histologic subtypes (*e.g.* serous, mucinous, endometrioid, and clear cell subtypes, as well as subclassifications and alternate classifications adenocarcinoma, papillary adenocarcinoma, papillary cystadenocarcinoma, surface papillary carcinoma, malignant adenofibroma, cystadenofibroma, adenocarcinoma, cystadenocarcinoma, adenoacanthoma,

endometrioid stromal sarcoma, mesodermal (Müllerian) mixed tumor, mesonephroid tumor, malignant carcinoma, Brenner tumor, mixed epithelial tumor, and undifferentiated carcinoma, using the WHO/FIGO system for classification of malignant breast tumors; Scully, *Atlas of Tumor Pathology*, 3d series, Washington DC), and  
5 various grades (*i.e.* grade I {well differentiated} , grade II {moderately well differentiated}, and grade III {poorly differentiated from surrounding normal tissue} ). In addition, as a greater number of patient samples are assessed for expression of the markers of the invention and the outcomes of the individual patients from whom the samples were obtained are correlated, it will also be confirmed that altered expression of  
10 certain of the markers of the invention are strongly correlated with malignant cancers and that altered expression of other markers of the invention are strongly correlated with benign tumors. The compositions, kits, and methods of the invention are thus useful for characterizing one or more of the stage, grade, histological type, and benign/malignant nature of breast cancer in patients. In addition, these compositions, kits, and methods  
15 can be used to detect and differentiate lobular and ductal carcinoma breast cancers.

When the compositions, kits, and methods of the invention are used for characterizing one or more of the stage, grade, histological type, and benign/malignant nature of breast cancer in a patient, it is preferred that the marker or panel of markers of the invention is selected such that a positive result is obtained in at least about 20%, and  
20 preferably at least about 40%, 60%, or 80%, and more preferably in substantially all patients afflicted with an breast cancer of the corresponding stage, grade, histological type, or benign/malignant nature. Preferably, the marker or panel of markers of the invention is selected such that a PPV of greater than about 10% is obtained for the general population (more preferably coupled with an assay specificity greater than  
25 99.5%).

When a plurality of markers of the invention are used in the compositions, kits, and methods of the invention, the level of expression of each marker in a patient sample can be compared with the normal level of expression of each of the plurality of markers in non-cancerous samples of the same type, either in a single reaction mixture (*i.e.* using  
30 reagents, such as different fluorescent probes, for each marker) or in individual reaction mixtures corresponding to one or more of the markers. In one embodiment, a significantly enhanced level of expression of more than one of the plurality of markers

in the sample, relative to the corresponding normal levels, is an indication that the patient is afflicted with breast cancer. In another embodiment, a significantly lower level of expression in the sample of each of the plurality of markers, relative to the corresponding normal levels, is an indication that the patient is afflicted with breast cancer. In yet another embodiment, a significantly enhanced level of expression of one or more markers and a significantly lower level of expression of one or more markers in a sample relative to the corresponding normal levels, is an indication that the patient is afflicted with breast cancer. When a plurality of markers is used, it is preferred that 2, 3, 4, 5, 8, 10, 12, 15, 20, 30, or 50 or more individual markers be used, wherein fewer markers are preferred.

In order to maximize the sensitivity of the compositions, kits, and methods of the invention (*i.e.* by interference attributable to cells of non-breast origin in a patient sample), it is preferable that the marker of the invention used therein be a marker which has a restricted tissue distribution, *e.g.*, normally not expressed in a non-breast tissue.

Only a small number of markers are known to be associated with breast cancers (*e.g.* *BRCA1* and *BRCA2*). These markers are not, of course, included among the markers of the invention, although they may be used together with one or more markers of the invention in a panel of markers, for example. It is well known that certain types of genes, such as oncogenes, tumor suppressor genes, growth factor-like genes, protease-like genes, and protein kinase-like genes are often involved with development of cancers of various types. Thus, among the markers of the invention, use of those which correspond to proteins which resemble known proteins encoded by known oncogenes and tumor suppressor genes, and those which correspond to proteins which resemble growth factors, proteases, and protein kinases are preferred.

Known oncogenes and tumor suppressor genes include, for example, *abl*, *abr*, *akt2*, *apc*, *bcl2 $\alpha$* , *bcl2 $\beta$* , *bcl3*, *bcr*, *brca1*, *brca2*, *cbl*, *ccnd1*, *cdc42*, *cdk4*, *crk- II*, *csf1r/fms*, *dbl*, *dcc*, *dpc4/smad4*, *e-cad*, *e2f1/rbap*, *egfr/erbB-1*, *elk1*, *elk3*, *eph*, *erg*, *ets1*, *ets2*, *fer*, *fgr/src2*, *fli1/ergb2*, *fos*, *fps/fes*, *fra1*, *fra2*, *fyn*, *hck*, *hek*, *her2/erbB- 2/neu*, *her3/erbB-3*, *her4/erbB-4*, *hras1*, *hst2*, *hstf1*, *igfbp2*, *ink4a*, *ink4b*, *int2/lfgf3*, *jun*, *junb*, *jund*, *kip2*, *kit*, *kras2a*, *kras2b*, *lck*, *lyn*, *mas*, *max*, *mcc*, *mdm2*, *met*, *mlh1*, *mmp10*, *mos*, *msh2*, *msh3*, *msh6*, *myb*, *myba*, *mybb*, *myc*, *mycl1*, *mycn*, *nf1*, *nf2*, *nme2*, *nras*, *p53*, *pdgfb*, *phb*, *pim1*, *pms1*, *pms2*, *ptc*, *pten*, *raf1*, *rap1a*, *rb1*, *rel*, *ret*, *ros1*, *ski*, *src1*, *tall*,

*tgfbr2, tgfb3, tgfbr3, thral, thrb, tiam1, timp3, tjp1, tp53, trk, vav, vhl, vil2, waf1, wnt1, wnt2, wt1, and yes1* (Hesketh, 1997, In: *The Oncogene and Tumour Suppressor Gene Facts Book*, 2nd Ed., Academic Press; Fishel *et al.*, 1994, *Science* 266:1403-1405).

- Known growth factors include platelet-derived growth factor alpha, platelet-
- 5 derived growth factor beta (simian sarcoma viral {v-sis} oncogene homolog), thrombopoietin (myeloproliferative leukemia virus oncogene ligand, megakaryocyte growth and development factor), erythropoietin, B cell growth factor, macrophage stimulating factor 1 (hepatocyte growth factor-like protein), hepatocyte growth factor (hepapoietin A), insulin-like growth factor 1 (somatomedia C), hepatoma-derived
- 10 growth factor, amphiregulin (schwannoma-derived growth factor), bone morphogenetic proteins 1, 2, 3, 3 beta, and 4, bone morphogenetic protein 7 (osteogenic protein 1), bone morphogenetic protein 8 (osteogenic protein 2), connective tissue growth factor, connective tissue activation peptide 3, epidermal growth factor (EGF), teratocarcinoma-derived growth factor 1, endothelin, endothelin 2, endothelin 3, stromal cell-derived
- 15 factor 1, vascular endothelial growth factor (VEGF), VEGF-B, VEGF-C, placental growth factor (vascular endothelial growth factor-related protein), transforming growth factor alpha, transforming growth factor beta 1 and its precursors, transforming growth factor beta 2 and its precursors, fibroblast growth factor 1 (acidic), fibroblast growth factor 2 (basic), fibroblast growth factor 5 and its precursors, fibroblast growth factor 6
- 20 and its precursors, fibroblast growth factor 7 (keratinocyte growth factor), fibroblast growth factor 8 (androgen-induced), fibroblast growth factor 9 (glia-activating factor), pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1), brain-derived neurotrophic factor, and recombinant glial growth factor 2.

- Known proteases include interleukin-1 beta convertase and its precursors, Mch6
- 25 and its precursors, Mch2 isoform alpha, Mch4, Cpp32 isoform alpha, Lice2 gamma cysteine protease, Ich-1S, Ich-1L, Ich-2 and its precursors, TY protease, matrix metalloproteinase 1 (interstitial collagenase), matrix metalloproteinase 2 (gelatinase A, 72kD gelatinase, 72kD type IV collagenase), matrix metalloproteinase 7 (matrilysin), matrix metalloproteinase 8 (neutrophil collagenase), matrix metalloproteinase 12
- 30 (macrophage elastase), matrix metalloproteinase 13 (collagenase 3), metalloproteinase 1, cysteine-rich metalloproteinase (disintegrin) and its precursors, subtilisin-like protease Pc8 and its precursors, chymotrypsin, snake venom-like protease, cathepsin 1, cathepsin D

(lysosomal aspartyl protease), stromelysin, aminopeptidase N, plasminogen, tissue plasminogen activator, plasminogen activator inhibitor type II, and urokinase-type plasminogen activator.

- Known protein kinases include DAP kinase, serine/threonine protein kinases
- 5 NIK, PK428, Krs-2, SAK, and EMK, interferon-inducible double stranded RNA dependent protein kinase, FAST kinase, AIM1, IPL1-like midbody-associated protein kinase-1, NIMA-like protein kinase 1 (NLK1), the cyclin-dependent kinases (cdk1-10), checkpoint kinase Chk1, Nek3 protein kinase, BMK1 beta kinase, Clk1, Clk2, Clk3, extracellular signal-regulated kinases 1, 3, and 6, cdc28 protein kinase 1, cdc28 protein
  - 10 kinase 2, pLK, Myt1, c-Jun N-terminal kinase 2, Cam kinase 1, the MAP kinases, insulin-stimulated protein kinase 1, beta-adrenergic receptor kinase 2, ribosomal protein S6 kinase, kinase suppressor of ras-1 (KSR1), putative serine/threonine protein kinase Prk, PkB kinase, cAMP-dependent protein kinase, cGMP-dependent protein kinase, type II cGMP-dependent protein kinase, protein kinases Dyrk2, Dyrk3, and Dyrk4, Rho-
  - 15 associated coiled-coil containing protein kinase p160ROCK, protein tyrosine kinase t-Ror1, Ste20-related kinases, cell adhesion kinase beta, protein kinase 3, stress-activated protein kinase 4, protein kinase Zpk, serine kinase hPAK65, dual specificity mitogen-activated protein kinases 1 and 2, casein kinase I gamma 2, p21-activated protein kinase Pak1, lipid-activated protein kinase PRK2, focal adhesion kinase, dual-specificity
  - 20 tyrosine-phosphorylation regulated kinase, myosin light chain kinase, serine kinases SRPK2, TESK1, and VRK2, B lymphocyte serine/threonine protein kinase, stress-activated protein kinases JNK1 and JNK2, phosphorylase kinase, protein tyrosine kinase Tec, Jak2 kinase, protein kinase Ndr, MEK kinase 3, SHB adaptor protein (a Src homology 2 protein), agammaglobulinaemia protein-tyrosine kinase (Atk), protein
  - 25 kinase ATR, guanylate kinase 1, thrombopoietin receptor and its precursors, DAG kinase epsilon, and kinases encoded by oncogenes or viral oncogenes such as v-fgr (Gardner-Rasheed), v-abl (Abelson murine leukemia viral oncogene homolog 1), v-arg (Abelson murine leukemia viral oncogene homolog, Abelson-related gene), v-fes and v-fps (feline sarcoma viral oncogene and Fujinami avian sarcoma viral oncogene
  - 30 homologs), proto-oncogene *c-cot*, oncogene *pim-1*, and oncogene *mas1*.

It is recognized that the compositions, kits, and methods of the invention will be of particular utility to patients having an enhanced risk of developing breast cancer and their medical advisors. Patients recognized as having an enhanced risk of developing breast cancer include, for example, patients having a familial history of breast cancer, 5 patients identified as having a mutant oncogene (*i.e.* at least one allele), and patients of advancing age (*i.e.* women older than about 50 or 60 years).

The level of expression of a marker in normal (*i.e.* non-cancerous) human breast tissue can be assessed in a variety of ways. In one embodiment, this normal level of expression is assessed by assessing the level of expression of the marker in a portion of 10 breast cells which appears to be non-cancerous and by comparing this normal level of expression with the level of expression in a portion of the breast cells which is suspected of being cancerous. For example, when mammography or other medical procedure, reveals the presence of a lump in a patient's breast, the normal level of expression of a marker may be assessed using the non-affected breast tissue, and this normal level of 15 expression may be compared with the level of expression of the same marker in an affected portion (*i.e.* the lump) of the affected breast. Alternately, and particularly as further information becomes available as a result of routine performance of the methods described herein, population-average values for normal expression of the markers of the invention may be used. In other embodiments, the 'normal' level of expression of a 20 marker may be determined by assessing expression of the marker in a patient sample obtained from a non-cancer-afflicted patient, from a patient sample obtained from a patient before the suspected onset of breast cancer in the patient, from archived patient samples, and the like.

The invention includes compositions, kits, and methods for assessing the presence 25 of breast cancer cells in a sample (*e.g.* an archived tissue sample or a sample obtained from a patient). These compositions, kits, and methods are substantially the same as those described above, except that, where necessary, the compositions, kits, and methods are adapted for use with samples other than patient samples. For example, when the sample to be used is a paraffinized, archived human tissue sample, it can be 30 necessary to adjust the ratio of compounds in the compositions of the invention, in the kits of the invention, or the methods used to assess levels of marker expression in the



sample. Such methods are well known in the art and within the skill of the ordinary artisan.

The invention includes a kit for assessing the presence of breast cancer cells (*e.g.* in a sample such as a patient sample). The kit comprises a plurality of reagents, each of which is capable of binding specifically with a nucleic acid or polypeptide corresponding to a marker of the invention. Suitable reagents for binding with a polypeptide corresponding to a marker of the invention include antibodies, antibody derivatives, antibody fragments, and the like. Suitable reagents for binding with a nucleic acid (*e.g.* a genomic DNA, an mRNA, a spliced mRNA, a cDNA, or the like) include complementary nucleic acids. For example, the nucleic acid reagents may include oligonucleotides (labeled or non-labeled) fixed to a substrate, labeled oligonucleotides not bound with a substrate, pairs of PCR primers, molecular beacon probes, and the like.

The kit of the invention may optionally comprise additional components useful for performing the methods of the invention. By way of example, the kit may comprise fluids (*e.g.* SSC buffer) suitable for annealing complementary nucleic acids or for binding an antibody with a protein with which it specifically binds, one or more sample compartments, an instructional material which describes performance of a method of the invention, a sample of normal breast cells, a sample of breast cancer cells, and the like.

The invention also includes a method of making an isolated hybridoma which produces an antibody useful for assessing whether patient is afflicted with breast cancer. In this method, a protein corresponding to a marker of the invention is isolated (*e.g.* by purification from a cell in which it is expressed or by transcription and translation of a nucleic acid encoding the protein *in vivo* or *in vitro* using known methods). A vertebrate, preferably a mammal such as a mouse, rat, rabbit, or sheep, is immunized using the isolated protein or protein fragment. The vertebrate may optionally (and preferably) be immunized at least one additional time with the isolated protein or protein fragment, so that the vertebrate exhibits a robust immune response to the protein or protein fragment. Splenocytes are isolated from the immunized vertebrate and fused with an immortalized cell line to form hybridomas, using any of a variety of methods well known in the art. Hybridomas formed in this manner are then screened using standard methods to identify one or more hybridomas which produce an antibody which

specifically binds with the protein or protein fragment. The invention also includes hybridomas made by this method and antibodies made using such hybridomas.

The invention also includes a method of assessing the efficacy of a test compound for inhibiting breast cancer cells. As described above, differences in the level of  
5 expression of the markers of the invention correlate with the cancerous state of breast cells. Although it is recognized that changes in the levels of expression of certain of the markers of the invention likely result from the cancerous state of breast cells, it is likewise recognized that changes in the levels of expression of other of the markers of the invention induce, maintain, and promote the cancerous state of those cells. Thus,  
10 compounds which inhibit breast cancer in a patient will cause the level of expression of one or more of the markers of the invention to change to a level nearer the normal level of expression for that marker (*i.e.* the level of expression for the marker in non-cancerous breast cells).

This method thus comprises comparing expression of a marker in a first breast  
15 cell sample and maintained in the presence of the test compound and expression of the marker in a second breast cell sample and maintained in the absence of the test compound. A significant alteration in the level of expression of a marker listed in Tables 1-6, may be an indication that the test compound inhibits breast cancer (*e.g.*, decreases in expression in those markers that are over-expressed in breast cancer cells or  
20 more aggressive breast cancer cells and breast cancer cells from patients with poor clinical outcome or increases expression in those markers that are under-expressed in breast cancer cells or in more aggressive breast cancer cells or breast cancer cells from patients with poor clinical outcome. The breast cell samples may, for example, be aliquots of a single sample of normal breast cells obtained from a patient, pooled  
25 samples of normal breast cells obtained from a patient, cells of a normal breast cell line, aliquots of a single sample of breast cancer cells obtained from a patient, pooled samples of breast cancer cells obtained from a patient, cells of a breast cancer cell line, or the like. In one embodiment, the samples are breast cancer cells obtained from a patient and a plurality of compounds known to be effective for inhibiting various breast cancers are  
30 tested in order to identify the compound which is likely to best inhibit the breast cancer in the patient.

This method may likewise be used to assess the efficacy of a therapy for inhibiting breast cancer in a patient. In this method, the level of expression of one or more markers of the invention in a pair of samples (one subjected to the therapy, the other not subjected to the therapy) is assessed. As with the method of assessing the efficacy of test compounds, if the therapy induces a significant alteration in the level of expression of a marker listed in Tables 1-6, or blocks induction of a marker listed in Tables 1-6, then the therapy may be efficacious for inhibiting breast cancer. As above, if samples from a selected patient are used in this method, then alternative therapies can be assessed *in vitro* in order to select a therapy most likely to be efficacious for inhibiting breast cancer in the patient.

As described herein, breast cancer in patients is associated with levels of expression of one or more markers listed in Tables 1-6. While, as discussed above, some of these changes in expression level result from occurrence of the breast cancer, others of these changes induce, maintain, and promote the cancerous state of breast cancer cells. Thus, breast cancer characterized by an alteration in the level of expression of one or more markers listed in Tables 1-6 can be inhibited by hampering or increasing expression of those markers.

Expression of a marker listed in Tables 1-6 can be inhibited in a number of ways generally known in the art. For example, an antisense oligonucleotide can be provided to the breast cancer cells in order to inhibit transcription, translation, or both, of the marker(s). Alternately, a polynucleotide encoding an antibody, an antibody derivative, or an antibody fragment, and operably linked with an appropriate promoter/regulator region, can be provided to the cell in order to generate intracellular antibodies which will inhibit the function or activity of the protein corresponding to the marker(s). Using the methods described herein, a variety of molecules, particularly including molecules sufficiently small that they are able to cross the cell membrane, can be screened in order to identify molecules which inhibit expression of the marker(s). The compound so identified can be provided to the patient in order to inhibit expression of the marker(s) in the breast cancer cells of the patient.

Expression of a marker listed within Tables 1-6 can be enhanced in number of ways generally known in the art. For example, a polynucleotide encoding the marker and operably linked with an appropriate promoter/regulator region can be provided to

- 27 -

breast cancer cells of the patient in order to induce enhanced expression of the protein (and mRNA) corresponding to the marker therein. Alternatively, if the protein is capable of crossing the cell membrane, inserting itself in the cell membrane, or is normally a secreted protein, then expression of the protein can be enhanced by providing the protein (*e.g.* directly or by way of the bloodstream or another breast-associated fluid) to breast cancer cells in the patient.

As described above, the cancerous state of human breast cells is correlated with changes in the levels of expression of the markers of the invention. The invention thus includes a method for assessing the human breast cell carcinogenic potential of a test compound. This method comprises maintaining separate aliquots of human breast cells in the presence and absence of the test compound. Expression of a marker of the invention in each of the aliquots is compared. A significant alteration in the level of expression of a marker listed in Tables 1-6 in the aliquot maintained in the presence of the test compound (relative to the aliquot maintained in the absence of the test compound) may be an indication that the test compound possesses human breast cell carcinogenic potential. The relative carcinogenic potentials of various test compounds can be assessed by comparing the degree of enhancement or inhibition of the level of expression of the relevant markers, by comparing the number of markers for which the level of expression is enhanced or inhibited, or by comparing both.

Various aspects of the invention are described in further detail in the following subsections.

#### I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to novel isolated nucleic acid molecules that correspond to a marker of the invention, including nucleic acids which encode a polypeptide corresponding to a marker of the invention or a portion of such a polypeptide. Isolated nucleic acids of the invention also include nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules that correspond to a marker of the invention, including nucleic acids which encode a polypeptide corresponding to a marker of the invention, and fragments of such nucleic acid molecules, *e.g.*, those suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is

intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

5       An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein-encoding sequences) which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the  
10       nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material,  
15       or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

      A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid encoding a protein corresponding to a marker listed in Tables 1-6, can be isolated using standard molecular biology techniques and the sequence information in the database records  
20       described herein. Using all or a portion of such nucleic acid sequences, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, ed., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

25       A process for identifying a larger fragment or the full-length coding sequence of a marker of the present invention is thus also provided. Any conventional recombinant DNA techniques applicable for isolating polynucleotides may also be employed. One such method involves the 5'-RACE-PCR technique, in which the poly-A mRNA that contains the coding sequence of particular interest is first reverse transcribed with a 3'-  
30       primer comprising a sequence disclosed herein. The newly synthesized cDNA strand is then tagged with an anchor primer with a known sequence, which preferably contains a convenient cloning restriction site attached at the 5' end. The tagged cDNA is then

- 29 -

amplified with the 3'-primer (or a nested primer sharing sequence homology to the internal sequences of the coding region) and the 5'-anchor primer. The amplification may be conducted under conditions of various levels of stringency to optimize the amplification specificity. 5'-RACE-PCR can be readily performed using commercial kits  
5 (available from, *e.g.*, BRL Life Technologies Inc., Clontech) according to the manufacturer's instructions.

Isolating the complete coding sequence of a gene can also be carried out in a hybridization assay using a suitable probe. The probe preferably comprises at least 10 nucleotides, and more preferably exhibits sequence homology to the polynucleotides of  
10 the markers of the present invention. Other high throughput screens for cDNAs, such as those involving gene chip technology, can also be employed in obtaining the complete cDNA sequence.

In addition, databases exist that reduce the complexity of ESTs by assembling contiguous EST sequences into tentative genes. For example, TIGR has assembled  
15 human ESTs into a database called THC for tentative human consensus sequences. The THC database allows for a more definitive assignment compared to ESTs alone. Software programs exist (TIGR assembler and TIGEM EST assembly machine and contig assembly program (see Huang, X. , 1996, *Genomes* 33:21-23)) that allow for assembling ESTs into contiguous sequences from any organism.

20 Alternatively, mRNA from a sample preparation is used to construct cDNA library in the ZAP Express vector following the procedure described in Velculescu *et al.*, 1997, *Science* 270:484. The ZAP Express cDNA synthesis kit (Stratagene) is used accordingly to the manufacturer's protocol. Plates containing 250 to 2000 plaques are hybridized as described in Rupert *et al.*, 1988, *Mol. Cell. Bio.* 8:3104 to oligonucleotide  
25 probes with the same conditions previously described for standard probes except that the hybridization temperature is reduced to a room temperature. Washes are performed in 6X standard-saline-citrate 0.1% SDS for 30 minutes at room temperature. The probes are labeled with <sup>32</sup>P-ATP through use of T4 polynucleotide kinase.

A partial cDNA (3' fragment) can be isolated by 3' directed PCR reaction. This  
30 procedure is a modification of the protocol described in Polyak *et al.*, 1997, *Nature* 389:300. Briefly, the procedure uses SAGE tags in PCR reaction such that the resultant PCR product contains the SAGE tag of interest as well as additional cDNA, the length

- 30 -

of which is defined by the position of the tag with respect to the 3' end of the cDNA. The cDNA product derived from such a transcript driven PCR reaction can be used for many applications.

RNA from a source to express the cDNA corresponding to a given tag is first  
5 converted to double-stranded cDNA using any standard cDNA protocol. Similar conditions used to generate cDNA for SAGE library construction can be employed except that a modified oligo-dT primer is used to derive the first strand synthesis. For example, the oligonucleotide of composition 5'-B-TCC GGC GCG CCG TTT TCC CAG TCA CGA(30)- 3', contains a poly-T stretch at the 3' end for hybridization and  
10 priming from poly-A tails, an M13 priming site for use in subsequent PCR steps, a 5' Biotin label (B) for capture to streptavidin-coated magnetic beads, and an AscI restriction endonuclease site for releasing the cDNA from the streptavidin-coated magnetic beads. Theoretically, any sufficiently-sized DNA region capable of hybridizing to a PCR primer can be used as well as any other 8 base pair recognizing endonuclease.

15 cDNA constructed utilizing this or similar modified oligo-dT primer is then processed exactly as described in U.S. Patent No. 5,695,937 up until adapter ligation where only one adapter is ligated to the cDNA pool. After Adapter ligation, the cDNA is released from the streptavidin-coated magnetic beads and is then used as a template for cDNA amplification.

20 Various PCR protocols can be employed using PCR priming sites within the 3' modified oligo-dT primer and the SAGE tag. The SAGE tag-derived PCR primer employed can be of varying length dictated by 5' extension of the tag into the adaptor sequence. cDNA products are now available for a variety of applications.

This technique can be further modified by: (1) altering the length and/or content  
25 of the modified oligo-dT primer; (2) ligating adaptors other than that previously employed within the SAGE protocol; (3) performing PCR from template retained on the streptavidin-coated magnetic beads; and (4) priming first strand cDNA synthesis with non-oligo-dT based primers.

Gene trapper technology can also be used. The reagents and manufacturer's  
30 instructions for this technology are commercially available from Life Technologies, Inc., Gaithersburg, Maryland. Briefly, a complex population of single-stranded phagemid DNA containing directional cDNA inserts is enriched for the target sequence by

- 31 -

hybridization in solution to a biotinylated oligonucleotide probe complementary to the target sequence. The hybrids are captured on streptavidin-coated paramagnetic beads. A magnet retrieves the paramagnetic beads from the solution, leaving nonhybridized single-stranded DNAs behind. Subsequently, the captured single-stranded DNA target is released from the biotinylated oligonucleotide. After release, the cDNA clone is further enriched by using a nonbiotinylated target oligonucleotide to specifically prime conversion of the single-stranded DNA. Following transformation and plating, typically 20% to 100% of the colonies represent the cDNA clone of interest. To identify the desired cDNA clone, the colonies may be screened by colony hybridization using the <sup>32</sup>P-labeled oligonucleotide as described above for solution hybridization, or alternatively by DNA sequencing and alignment of all sequences obtained from numerous clones to determine a consensus sequence.

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA, or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which has a nucleotide sequence complementary to the nucleotide sequence of a nucleic acid corresponding to a marker of the invention or to the nucleotide sequence of a nucleic acid encoding a protein which corresponds to a marker of the invention. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence, wherein the full length nucleic acid sequence comprises a marker of the invention or which encodes a polypeptide corresponding to a marker of the invention. Such nucleic acids can be used, for example, as a probe or primer. The probe/primer typically is used as one or more substantially purified oligonucleotides.



- 32 -

The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, preferably about 15, more preferably about 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, or 400 or more consecutive nucleotides of a nucleic acid of the invention.

5 Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences corresponding to one or more markers of the invention. The probe comprises a label group attached thereto, *e.g.*, a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-  
10 express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, *e.g.*, detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

The invention further encompasses nucleic acid molecules that differ, due to degeneracy of the genetic code, from the nucleotide sequence of nucleic acids encoding  
15 a protein which corresponds to a marker of the invention, and thus encode the same protein.

In addition to the nucleotide sequences described herein, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence can exist within a population (*e.g.*, the human population). Such  
20 genetic polymorphisms can exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. In addition, it will be appreciated that DNA polymorphisms that affect RNA expression levels can also exist that may affect the overall expression level of that gene (*e.g.*, by affecting regulation or degradation).

25 As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide corresponding to a marker of the invention. Such natural allelic variations can typically result in 1-5%  
30 variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a

variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 7, 15, 20, 25, 30, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid corresponding to a marker of the invention or to a nucleic acid encoding a protein corresponding to a marker of the invention. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 75% (80%, 85%, preferably 90%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in sections 6.3.1-6.3.6 of *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989). A preferred, non-limiting example of stringent hybridization conditions for annealing two single-stranded DNA each of which is at least about 100 bases in length and/or for annealing a single-stranded DNA and a single-stranded RNA each of which is at least about 100 bases in length, are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50-65°C. Further preferred hybridization conditions are taught in Lockhart, *et al.*, *Nature Biotechnology*, Volume 14, 1996 August:1675-1680; Breslauer, *et al.*, *Proc. Natl. Acad. Sci. USA*, Volume 83, 1986 June: 3746-3750; Van Ness, *et al.*, *Nucleic Acids Research*, Volume 19, No. 19, 1991 September: 5143-5151; McGraw, *et al.*, *BioTechniques*, Volume 8, No. 6 1990: 674-678; and Milner, *et al.*, *Nature Biotechnology*, Volume 15, 1997 June: 537-541, all expressly incorporated by reference.

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention that can exist in the population, the skilled artisan will further appreciate that sequence changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein encoded thereby. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type

sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologs of various species (*e.g.*, murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from the naturally-occurring proteins which correspond to the markers of the invention, yet retain biological activity. In one embodiment, such a protein has an amino acid sequence that is at least about 40% identical, 50%, 60%, 70%, 80%, 90%, 95%, or 98% identical to the amino acid sequence of one of the proteins which correspond to the markers of the invention.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of nucleic acids of the invention, such that one or more amino acid residue substitutions, additions, or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis,

- 35 -

and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

The present invention encompasses antisense nucleic acid molecules, *i.e.*,  
5 molecules which are complementary to a sense nucleic acid of the invention, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule corresponding to a marker of the invention or complementary to an mRNA sequence corresponding to a marker of the invention. Accordingly, an antisense nucleic acid of the invention can hydrogen bond to (*i.e.* anneal with) a sense nucleic acid of the  
10 invention. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can also be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3'  
15 sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense  
20 oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be  
25 used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-  
30 methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-

- 36 -

N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

5 Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been sub-cloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

10 The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a polypeptide corresponding to a selected marker of the invention to thereby inhibit expression of the marker, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide  
15 complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. Examples of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site or infusion of the antisense nucleic acid into an breast-associated body fluid. Alternatively,  
20 antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense  
25 nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an  $\alpha$ -anomeric nucleic  
30 acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\alpha$ -units, the strands run parallel to each other (Gaultier *et al.*, 1987, *Nucleic Acids Res.* 15:6625-6641). The

antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.*, 1987, *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987, *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA  
5 molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach, 1988, *Nature* 334:585-591) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity  
10 for a nucleic acid molecule encoding a polypeptide corresponding to a marker of the invention can be designed based upon the nucleotide sequence of a cDNA corresponding to the marker. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved (see Cech *et al.* U.S. Patent No.  
15 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742). Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see, *e.g.*, Bartel and Szostak, 1993, *Science* 261:1411-1418).

The invention also encompasses nucleic acid molecules which form triple helical  
20 structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. See generally Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-  
25 36; and Maher (1992) *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic  
30 acids (see Hyrup *et al.*, 1996, *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a

pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup (1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup, 1996, *supra*; Perry-O'Keefe *et al.*, 1996, *Proc. Natl. Acad. Sci. USA* 93:14670-675).

In another embodiment, PNAs can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which can combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNASE H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup, 1996, *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, and Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag *et al.*, 1989, *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a step-wise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.*, 1996, *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be

synthesized with a 5' DNA segment and a 3' PNA segment (Peterser *et al.*, 1975, *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide can include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.*, 1988, *Bio/Techniques* 6:958-976) or intercalating agents (see, *e.g.*, Zon, 1988, *Pharm. Res.* 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The invention also includes molecular beacon nucleic acids having at least one region which is complementary to a nucleic acid of the invention, such that the molecular beacon is useful for quantitating the presence of the nucleic acid of the invention in a sample. A "molecular beacon" nucleic acid is a nucleic acid comprising a pair of complementary regions and having a fluorophore and a fluorescent quencher associated therewith. The fluorophore and quencher are associated with different portions of the nucleic acid in such an orientation that when the complementary regions are annealed with one another, fluorescence of the fluorophore is quenched by the quencher. When the complementary regions of the nucleic acid are not annealed with one another, fluorescence of the fluorophore is quenched to a lesser degree. Molecular beacon nucleic acids are described, for example, in U.S. Patent 5,876,930.

## II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins which correspond to individual markers of the invention, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide corresponding to a marker of the invention. In one embodiment, the native polypeptide corresponding to a marker can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification



techniques. In another embodiment, polypeptides corresponding to a marker of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide corresponding to a marker of the invention can be synthesized chemically using standard peptide synthesis techniques.

5       An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the  
10   protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is  
15   also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, *i.e.*, it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such  
20   preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide corresponding to a marker of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein corresponding to the marker,  
25   which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other  
30   biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have amino acid sequences encoded by the nucleic acid sequences described herein. Other useful proteins are substantially identical (*e.g.*, at least about 40%, preferably 50%, 60%, 70%, 80%, 90%, 95%, or 99%) to one of these sequences and retain the functional activity of the protein of the corresponding  
5 naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal  
10 alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences  
15 is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions)  $\times 100$ ). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a  
20 mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with  
25 the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in  
30 Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the

- 42 -

default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the

5 ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Yet another useful algorithm for identifying regions of local sequence similarity and alignment is the FASTA algorithm as described in Pearson and Lipman (1988)

10 *Proc. Natl. Acad. Sci. USA* 85:2444-2448. When using the FASTA algorithm for comparing nucleotide or amino acid sequences, a PAM120 weight residue table can, for example, be used with a *k*-tuple value of 2.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent

15 identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins corresponding to a marker of the invention. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably a biologically active part) of a polypeptide corresponding to a marker of the invention operably linked to a heterologous

20 polypeptide (*i.e.*, a polypeptide other than the polypeptide corresponding to the marker). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the amino-terminus or the carboxyl-terminus of the polypeptide of the invention.

25 One useful fusion protein is a GST fusion protein in which a polypeptide corresponding to a marker of the invention is fused to the carboxyl terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal

30 sequence at its amino terminus. For example, the native signal sequence of a polypeptide corresponding to a marker of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence

of the baculovirus envelope protein can be used as a heterologous signal sequence (Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, NY, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; 5 La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook *et al.*, *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide corresponding to a marker of the invention 10 is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion 15 protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction can be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a 20 polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. 25 Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see, *e.g.*, Ausubel *et al.*, *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). 30 A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain

5 processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal sequence, as well as to polypeptides from which the signal sequence has been proteolytically cleaved (*i.e.*, the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence can be operably linked

10 in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods.

15 Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

The present invention also pertains to variants of the polypeptides corresponding to individual markers of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants

20 can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which

25 includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

30 Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the protein of the invention for agonist or antagonist activity. In

one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, 1983, *Tetrahedron* 39:3; Itakura *et al.*, 1984, *Annu. Rev. Biochem.* 53:323; Itakura *et al.*, 1984, *Science* 198:1056; Ike *et al.*, 1983 *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide corresponding to a marker of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes amino terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional

mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan, 1992, *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.*, 1993, *Protein Engineering* 6(3):327- 331).

An isolated polypeptide corresponding to a marker of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30 or more) amino acid residues of the amino acid sequence of one of the polypeptides of the invention, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with a marker of the invention to which the protein corresponds. Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, *e.g.*, hydrophilic regions. Hydrophobicity sequence analysis, hydrophilicity sequence analysis, or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable (*i.e.* immunocompetent) subject such as a rabbit, goat, mouse, or other mammal or vertebrate. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed or chemically-synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or a similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The terms "antibody" and "antibody substance" as used interchangeably herein refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention, *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated

by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be harvested or isolated from the subject (*e.g.*, from the blood or serum of the subject) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. Alternatively, antibodies specific for a protein or polypeptide of the invention can be selected or (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than



- 48 -

those of the desired protein or polypeptide of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or

5 polypeptide of the invention.

At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B

10 cell hybridoma technique (see Kozbor *et al.*, 1983, *Immunol. Today* 4:72), the EBV-hybridoma technique (see Cole *et al.*, pp. 77-96 In *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., 1985) or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology*, Coligan *et al.* ed., John Wiley & Sons, New York, 1994). Hybridoma cells producing a

15 monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified

20 and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods

25 and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT

30 Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734.

- 49 -

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.*, Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397, which are incorporated herein by reference in their entirety.) Humanized antibodies are antibody molecules from non-human species having one or more complementarily determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, *e.g.*, Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

Antibodies of the invention may be used as therapeutic agents in treating cancers. In a preferred embodiment, completely human antibodies of the invention are used for therapeutic treatment of human cancer patients, particularly those having breast cancer. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chain genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide corresponding to a marker of the invention. Monoclonal antibodies directed

- 50 -

against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope (Jespers *et al.*, 1994, *Bio/technology* 12:899-903).

An antibody directed against a polypeptide corresponding to a marker of the invention (e.g., a monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the marker (e.g., in a cellular lysate or cell supernatant) in order to evaluate the level and pattern of expression of the marker. The antibodies can also be used diagnostically to monitor protein levels in tissues or body fluids (e.g. in an ovary-associated body fluid) as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate,

rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

- 5 Further, an antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, 10 dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, 15 thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic 20 agents (*e.g.*, vincristine and vinblastine).

- The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin 25 such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), 30 granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in *Controlled Drug Delivery* (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in *Monoclonal Antibodies '84: Biological And Clinical Applications*, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.*, 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

Accordingly, in one aspect, the invention provides substantially purified antibodies or fragments thereof, and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences of the present invention, an amino acid sequence encoded by the cDNA of the present invention, a fragment of at least 15 amino acid residues of an amino acid sequence of the present invention, an amino acid sequence which is at least 95% identical to the amino acid sequence of the present invention (wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4) and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleic acid molecules of the present invention, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of

the present invention, an amino acid sequence encoded by the cDNA of the present invention, a fragment of at least 15 amino acid residues of the amino acid sequence of the present invention, an amino acid sequence which is at least 95% identical to the amino acid sequence of the present invention (wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4) and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleic acid molecules of the present invention, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences of the present invention, an amino acid sequence encoded by the cDNA of the present invention, a fragment of at least 15 amino acid residues of an amino acid sequence of the present invention, an amino acid sequence which is at least 95% identical to an amino acid sequence of the present invention (wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4) and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleic acid molecules of the present invention, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The substantially purified antibodies or fragments thereof may specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain or cytoplasmic membrane of a polypeptide of the invention. In a particularly preferred embodiment, the substantially purified antibodies or fragments

thereof, the non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind to a secreted sequence or an extracellular domain of the amino acid sequences of the present invention.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety  
5 or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention  
10 conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

15 Still another aspect of the invention is a method of making an antibody that specifically recognizes a polypeptide of the present invention, the method comprising immunizing a mammal with a polypeptide. The polypeptide used as an immungen comprises an amino acid sequence selected from the group consisting of the amino acid sequence of the present invention, an amino acid sequence encoded by the cDNA of the  
20 nucleic acid molecules of the present invention, a fragment of at least 15 amino acid residues of the amino acid sequence of the present invention, an amino acid sequence which is at least 95% identical to the amino acid sequence of the present invention (wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a  
25 gap penalty of 4) and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleic acid molecules of the present invention, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C.

After immunization, a sample is collected from the mammal that contains an  
30 antibody that specifically recognizes the polypeptide. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, the antibodies can be further purified from the sample using techniques well known to those of skill in the art.

- 55 -

The method can further comprise producing a monoclonal antibody-producing cell from the cells of the mammal. Optionally, antibodies are collected from the antibody-producing cell.

### 5 III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide corresponding to a marker of the invention (or a portion of such a polypeptide). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has  
10 been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and  
15 episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, namely expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA  
20 techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of  
25 the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of  
30 interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory



sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Methods in Enzymology: Gene Expression Technology* vol.185, Academic Press, San Diego, CA (1991). Regulatory sequences include those which  
5 direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The  
10 expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide corresponding to a marker of the invention in prokaryotic  
15 (e.g., *E. coli*) or eukaryotic cells (e.g., insect cells {using baculovirus expression vectors}, yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

20 Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein;  
25 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes,  
30 and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988, *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and

pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, 1988, *Gene* 69:301-315) and pET 11d (Studier *et al.*, p. 60-89, In *Gene Expression Technology: Methods in Enzymology* vol.185, Academic Press, San Diego, CA, 1991). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a co-expressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, p. 119-128, In *Gene Expression Technology: Methods in Enzymology* vol. 185, Academic Press, San Diego, CA, 1990). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, 1992, *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.*, 1987, *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, 1982, *Cell* 30:933-943), pJRY88 (Schultz *et al.*, 1987, *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, 1983, *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers, 1989, *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987, *Nature* 329:840) and pMT2PC

(Kaufman *et al.*, 1987, *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both  
5 prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook *et al.*, *supra*.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-  
10 specific promoters include the albumin promoter (liver-specific; Pinkert *et al.*, 1987, *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton, 1988, *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989, *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.*, 1983, *Cell* 33:729-740; Queen and Baltimore, 1983, *Cell* 33:741-748), neuron-specific promoters  
15 (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989, *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.*, 1985, *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters  
20 (Kessel and Gruss, 1990, *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Camper and Tilghman, 1989, *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a  
25 manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or  
30 regulatory sequences can be chosen which direct constitutive, tissue-specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid, or attenuated virus in which antisense nucleic

acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, 1986, *Trends in Genetics*, Vol. 1(1).

5 Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due  
10 to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (*e.g.*, *E. coli*) or eukaryotic cell (*e.g.*, insect cells, yeast or mammalian cells).

15 Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection,  
20 lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these  
25 integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable  
30 marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide corresponding to a marker of the invention. Accordingly, the invention further provides methods for producing a polypeptide corresponding to a marker of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the marker is produced. In another embodiment, the method further comprises isolating the marker polypeptide from the medium or the host cell.

10 The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequences encoding a polypeptide corresponding to a marker of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences  
15 encoding a marker protein of the invention have been introduced into their genome or homologous recombinant animals in which endogenous gene(s) encoding a polypeptide corresponding to a marker of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide corresponding to the marker and for identifying and/or evaluating modulators of polypeptide activity. As  
20 used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal  
25 develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and  
30 an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a nucleic acid encoding a polypeptide corresponding to a marker of the invention into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide corresponding to a marker of the invention into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient

- 62 -

length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi, 1987, *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g.,  
5 by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (see, e.g., Li *et al.*, 1992, *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley, *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, Ed., IRL, Oxford, 1987, pp.  
10 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination  
15 vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication NOS. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One  
20 example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.*, 1991, *Science* 251:1351-1355). If a *cre/loxP* recombinase system is used to regulate expression of the  
25 transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

30 Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813 and PCT Publication NOS. WO 97/07668 and WO 97/07669.

#### IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") corresponding to a marker of the invention can be incorporated  
5 into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents,  
10 and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

15 The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid corresponding to a marker of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid corresponding to a marker of the invention. Such compositions can  
20 further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid corresponding to a marker of the invention and one or more additional active compounds.

25 The invention also provides methods (also referred to herein as "screening assays") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, peptoids, small molecules or other drugs) which (a) bind to the marker, or (b) have a modulatory (*e.g.*, stimulatory or inhibitory) effect on the activity of the marker or, more specifically, (c) have a modulatory effect on the  
30 interactions of the marker with one or more of its natural substrates (*e.g.*, peptide, protein, hormone, co-factor, or nucleic acid), or (d) have a modulatory effect on the expression of the marker. Such assays typically comprise a reaction between the marker



- 64 -

and one or more assay components. The other components may be either the test compound itself, or a combination of test compound and a natural binding partner of the marker.

The test compounds of the present invention may be obtained from any available  
5 source, including systematic libraries of natural and/or synthetic compounds. Test compounds may also be obtained by any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless  
10 remain bioactive; see, *e.g.*, Zuckermann *et al.*, 1994, *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other  
15 four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med.*  
20 *Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992, *Biotechniques* 13:412-421), or on beads (Lam, 1991, *Nature* 354:82-84), chips (Fodor,  
25 1993, *Nature* 364:555-556), bacteria and/or spores, (Ladner, USP 5,223,409), plasmids (Cull *et al.*, 1992, *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith, 1990, *Science* 249:386-390; Devlin, 1990, *Science* 249:404-406; Cwirla *et al.*, 1990, *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici, 1991, *J. Mol. Biol.* 222:301-310; Ladner, *supra.*).

30 In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a marker or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test

- 65 -

compounds which bind to a marker or biologically active portion thereof. Determining the ability of the test compound to directly bind to a marker can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to the marker can be determined by detecting the labeled  
5 marker compound in a complex. For example, compounds (e.g., marker substrates) can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, assay components can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by  
10 determination of conversion of an appropriate substrate to product.

In another embodiment, the invention provides assays for screening candidate or test compounds which modulate the activity of a marker or a biologically active portion thereof. In all likelihood, the marker can, *in vivo*, interact with one or more molecules, such as but not limited to, peptides, proteins, hormones, cofactors and nucleic acids. For  
15 the purposes of this discussion, such cellular and extracellular molecules are referred to herein as "binding partners" or marker "substrate".

One necessary embodiment of the invention in order to facilitate such screening is the use of the marker to identify its natural *in vivo* binding partners. There are many ways to accomplish this which are known to one skilled in the art. One example is the  
20 use of the marker protein as "bait protein" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al*, 1993, *Cell* 72:223-232; Madura *et al*, 1993, *J. Biol. Chem.* 268:12046-12054; Bartel *et al*, 1993, *Biotechniques* 14:920-924; Iwabuchi *et al*, 1993 *Oncogene* 8:1693-1696; Brent WO94/10300) in order to identify other proteins which bind to or interact with the marker (binding partners) and,  
25 therefore, are possibly involved in the natural function of the marker. Such marker binding partners are also likely to be involved in the propagation of signals by the marker or downstream elements of a marker-mediated signaling pathway. Alternatively, such marker binding partners may also be found to be inhibitors of the marker.

The two-hybrid system is based on the modular nature of most transcription  
30 factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that encodes a marker protein fused to a gene encoding the DNA binding domain of a known

- 66 -

transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a marker-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be readily detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the marker protein.

In a further embodiment, assays may be devised through the use of the invention for the purpose of identifying compounds which modulate (*e.g.*, affect either positively or negatively) interactions between a marker and its substrates and/or binding partners. Such compounds can include, but are not limited to, molecules such as antibodies, peptides, hormones, oligonucleotides, nucleic acids, and analogs thereof. Such compounds may also be obtained from any available source, including systematic libraries of natural and/or synthetic compounds. The preferred assay components for use in this embodiment is an breast cancer marker identified herein, the known binding partner and/or substrate of same, and the test compound. Test compounds can be supplied from any source.

The basic principle of the assay systems used to identify compounds that interfere with the interaction between the marker and its binding partner involves preparing a reaction mixture containing the marker and its binding partner under conditions and for a time sufficient to allow the two products to interact and bind, thus forming a complex. In order to test an agent for inhibitory activity, the reaction mixture is prepared in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the marker and its binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the marker and its binding partner is then detected. The formation of a complex in the control reaction, but less or no such formation in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the marker and its binding

- 67 -

partner. Conversely, the formation of more complex in the presence of compound than in the control reaction indicates that the compound may enhance interaction of the marker and its binding partner.

The assay for compounds that interfere with the interaction of the marker with its binding partner may be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the marker or its binding partner onto a solid phase and detecting complexes anchored to the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the markers and the binding partners (*e.g.*, by competition) can be identified by conducting the reaction in the presence of the test substance, *i.e.*, by adding the test substance to the reaction mixture prior to or simultaneously with the marker and its interactive binding partner. Alternatively, test compounds that disrupt preformed complexes, *e.g.*, compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

In a heterogeneous assay system, either the marker or its binding partner is anchored onto a solid surface or matrix, while the other corresponding non-anchored component may be labeled, either directly or indirectly. In practice, microtitre plates are often utilized for this approach. The anchored species can be immobilized by a number of methods, either non-covalent or covalent, that are typically well known to one who practices the art. Non-covalent attachment can often be accomplished simply by coating the solid surface with a solution of the marker or its binding partner and drying. Alternatively, an immobilized antibody specific for the assay component to be anchored can be used for this purpose. Such surfaces can often be prepared in advance and stored.

In related embodiments, a fusion protein can be provided which adds a domain that allows one or both of the assay components to be anchored to a matrix. For example, glutathione-S-transferase/marker fusion proteins or glutathione-S-transferase/binding partner can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then

combined with the test compound or the test compound and either the non-adsorbed marker or its binding partner, and the mixture incubated under conditions conducive to complex formation (*e.g.*, physiological conditions). Following incubation, the beads or microtiter plate wells are washed to remove any unbound assay components, the  
5 immobilized complex assessed either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of marker binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a marker or a marker binding  
10 partner can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated marker protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In certain embodiments, the protein-immobilized surfaces can be prepared in  
15 advance and stored.

In order to conduct the assay, the corresponding partner of the immobilized assay component is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted assay components are removed (*e.g.*, by washing) and any complexes formed will remain immobilized on the solid surface. The detection  
20 of complexes anchored on the solid surface can be accomplished in a number of ways. Where the non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the  
25 initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds which modulate (*inhibit* or *enhance*) complex formation or which disrupt preformed complexes can be detected.

In an alternate embodiment of the invention, a homogeneous assay may be used.  
30 This is typically a reaction, analogous to those mentioned above, which is conducted in a liquid phase in the presence or absence of the test compound. The formed complexes are then separated from unreacted components, and the amount of complex formed is

determined. As mentioned for heterogeneous assay systems, the order of addition of reactants to the liquid phase can yield information about which test compounds modulate (inhibit or enhance) complex formation and which disrupt preformed complexes.

5           In such a homogeneous assay, the reaction products may be separated from unreacted assay components by any of a number of standard techniques, including but not limited to: differential centrifugation, chromatography, electrophoresis and immunoprecipitation. In differential centrifugation, complexes of molecules may be separated from uncomplexed molecules through a series of centrifugal steps, due to the  
10 different sedimentation equilibria of complexes based on their different sizes and densities (see, for example, Rivas, G., and Minton, A.P., *Trends Biochem Sci* 1993 Aug;18(8):284-7). Standard chromatographic techniques may also be utilized to separate complexed molecules from uncomplexed ones. For example, gel filtration chromatography separates molecules based on size, and through the utilization of an  
15 appropriate gel filtration resin in a column format, for example, the relatively larger complex may be separated from the relatively smaller uncomplexed components. Similarly, the relatively different charge properties of the complex as compared to the uncomplexed molecules may be exploited to differentially separate the complex from the remaining individual reactants, for example through the use of ion-exchange  
20 chromatography resins. Such resins and chromatographic techniques are well known to one skilled in the art (see, e.g., Heegaard, 1998, *J Mol. Recognit.* 11:141-148; Hage and Tweed, 1997, *J. Chromatogr. B. Biomed. Sci. Appl.*, 699:499-525). Gel electrophoresis may also be employed to separate complexed molecules from unbound species (see, e.g., Ausubel *et al* (eds.), In: *Current Protocols in Molecular Biology*, J. Wiley & Sons,  
25 New York. 1999). In this technique, protein or nucleic acid complexes are separated based on size or charge, for example. In order to maintain the binding interaction during the electrophoretic process, non-denaturing gels in the absence of reducing agent are typically preferred, but conditions appropriate to the particular interactants will be well known to one skilled in the art. Immunoprecipitation is another common technique  
30 utilized for the isolation of a protein-protein complex from solution (see, e.g., Ausubel *et al* (eds.), In: *Current Protocols in Molecular Biology*, J. Wiley & Sons, New York. 1999). In this technique, all proteins binding to an antibody specific to one of the

- 70 -

binding molecules are precipitated from solution by conjugating the antibody to a polymer bead that may be readily collected by centrifugation. The bound assay components are released from the beads (through a specific proteolysis event or other technique well known in the art which will not disturb the protein-protein interaction in the complex), and a second immunoprecipitation step is performed, this time utilizing antibodies specific for the correspondingly different interacting assay component. In this manner, only formed complexes should remain attached to the beads. Variations in complex formation in both the presence and the absence of a test compound can be compared, thus offering information about the ability of the compound to modulate interactions between the marker and its binding partner.

Also within the scope of the present invention are methods for direct detection of interactions between the marker and its natural binding partner and/or a test compound in a homogeneous or heterogeneous assay system without further sample manipulation. For example, the technique of fluorescence energy transfer may be utilized (see, *e.g.*, Lakowicz *et al*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al*, U.S. Patent No. 4,868,103). Generally, this technique involves the addition of a fluorophore label on a first 'donor' molecule (*e.g.*, marker or test compound) such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule (*e.g.*, marker or test compound), which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, spatial relationships between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter). A test substance which either enhances or hinders participation of one of the species in the preformed complex will result in the generation of a signal variant to that of background. In this way, test substances that modulate interactions between a marker and its binding partner can be identified in controlled assays.

In another embodiment, modulators of marker expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA or protein, corresponding to a marker in the cell, is determined. The level of expression of mRNA or protein in the presence of the candidate compound is compared  
5 to the level of expression of mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of marker expression based on this comparison. For example, when expression of marker mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of  
10 marker mRNA or protein expression. Conversely, when expression of marker mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of marker mRNA or protein expression. The level of marker mRNA or protein expression in the cells can be determined by methods described herein for detecting marker mRNA  
15 or protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a marker protein can be further confirmed *in vivo*, *e.g.*, in a whole animal model for  
20 cellular transformation and/or tumorigenesis.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, an marker modulating agent, an antisense  
25 marker nucleic acid molecule, an marker-specific antibody, or an marker-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the  
30 above-described screening assays for treatments as described herein.



It is understood that appropriate doses of small molecule agents and protein or polypeptide agents depends upon a number of factors within the knowledge of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of these agents will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the agent to have upon the nucleic acid or polypeptide of the invention. Exemplary doses of a small molecule include milligram or microgram amounts per kilogram of subject or sample weight (*e.g.* about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram). Exemplary doses of a protein or polypeptide include gram, milligram or microgram amounts per kilogram of subject or sample weight (*e.g.* about 1 microgram per kilogram to about 5 grams per kilogram, about 100 micrograms per kilogram to about 500 milligrams per kilogram, or about 1 milligram per kilogram to about 50 milligrams per kilogram). It is furthermore understood that appropriate doses of one of these agents depend upon the potency of the agent with respect to the expression or activity to be modulated. Such appropriate doses can be determined using the assays described herein. When one or more of these agents is to be administered to an animal (*e.g.* a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher can, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific agent employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following

components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediamine-tetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound

into a sterile vehicle which contains a basic dispersion medium, and then incorporating the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active  
5 ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and  
10 used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches, and the like can  
15 contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as  
20 peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For  
25 transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active  
30 compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will  
5 protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art.  
10 The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes having monoclonal antibodies incorporated therein or thereon) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

15 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required  
20 pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight  
25 (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to  
30 enhance uptake and tissue penetration (*e.g.*, into the breast epithelium). A method for lipidation of antibodies is described by Cruikshank *et al.* (1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193.

- 76 -

The nucleic acid molecules corresponding to a marker of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470), or by stereotactic injection (see, *e.g.*, Chen *et al.*, 1994, *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.* retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### V. Electronic Apparatus Readable Media and Arrays

Electronic apparatus readable media comprising a breast cancer marker of the present invention is also provided. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon a marker of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the markers of the present invention.

5       A variety of software programs and formats can be used to store the marker information of the present invention on the electronic apparatus readable medium. For example, the nucleic acid sequence corresponding to the markers can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in  
10   a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of dataprocessor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the markers of the present invention.

By providing the markers of the invention in readable form, one can routinely  
15   access the marker sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the present invention in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular  
20   target sequence or target motif.

The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has breast cancer or a pre-disposition to breast cancer, wherein the method comprises the steps of determining the presence or absence of a breast cancer marker and based on the presence or absence of  
25   the breast cancer marker, determining whether the subject has breast cancer or a pre-disposition to breast cancer and/or recommending a particular treatment for the breast cancer or pre- breast cancer condition.

The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has breast cancer or a pre-  
30   disposition to breast cancer associated with a breast cancer marker wherein the method comprises the steps of determining the presence or absence of the breast cancer marker, and based on the presence or absence of the breast cancer marker, determining whether

the subject has breast cancer or a pre-disposition to breast cancer, and/or recommending a particular treatment for the breast cancer or pre- breast cancer condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the  
5 subject.

The present invention also provides in a network, a method for determining whether a subject has breast cancer or a pre-disposition to breast cancer associated with a breast cancer marker, said method comprising the steps of receiving information associated with the breast cancer marker receiving phenotypic information associated  
10 with the subject, acquiring information from the network corresponding to the breast cancer marker and/or breast cancer, and based on one or more of the phenotypic information, the breast cancer marker, and the acquired information, determining whether the subject has breast cancer or a pre-disposition to breast cancer. The method may further comprise the step of recommending a particular treatment for the breast  
15 cancer or pre- breast cancer condition.

The present invention also provides a business method for determining whether a subject has breast cancer or a pre-disposition to breast cancer, said method comprising the steps of receiving information associated with the breast cancer marker, receiving phenotypic information associated with the subject, acquiring information from the  
20 network corresponding to the breast cancer marker and/or breast cancer, and based on one or more of the phenotypic information, the breast cancer marker, and the acquired information, determining whether the subject has breast cancer or a pre-disposition to breast cancer. The method may further comprise the step of recommending a particular treatment for the breast cancer or pre- breast cancer condition.

25 The invention also includes an array comprising a breast cancer marker of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression. This allows a profile to be  
30 developed showing a battery of genes specifically expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of breast cancer, progression of breast cancer, and processes, such a cellular transformation associated with breast cancer.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells. This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes that could serve as a molecular target for diagnosis or therapeutic intervention.

## VI. Predictive Medicine

The present invention pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials



- 80 -

are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining the level of expression of polypeptides or nucleic acids corresponding to one or more markers of the invention, in order to determine whether  
5 an individual is at risk of developing breast cancer. Such assays can be used for prognostic or predictive purposes to thereby prophylactically treat an individual prior to the onset of the cancer.

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds administered either to inhibit breast cancer or to treat or  
10 prevent any other disorder {i.e. in order to understand any breast carcinogenic effects that such treatment may have} ) on the expression or activity of a marker of the invention in clinical trials. These and other agents are described in further detail in the following sections.

#### 15     A. Diagnostic Assays

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid corresponding to a marker of the invention in a biological sample involves obtaining a biological sample (e.g. a breast-associated body fluid) from a test subject and contacting the biological sample with a compound or an agent capable of detecting  
20 the polypeptide or nucleic acid (e.g., mRNA, genomic DNA, or cDNA). The detection methods of the invention can thus be used to detect mRNA, protein, cDNA, or genomic DNA, for example, in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide corresponding to a  
25 marker of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of a polypeptide corresponding to a marker of the invention include introducing into a subject a labeled antibody directed against the polypeptide.  
30 For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

A general principle of such diagnostic and prognostic assays involves preparing a sample or reaction mixture that may contain a marker, and a probe, under appropriate conditions and for a time sufficient to allow the marker and probe to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture.

5 These assays can be conducted in a variety of ways.

For example, one method to conduct such an assay would involve anchoring the marker or probe onto a solid phase support, also referred to as a substrate, and detecting target marker/probe complexes anchored on the solid phase at the end of the reaction.

In one embodiment of such a method, a sample from a subject, which is to be assayed  
10 for presence and/or concentration of marker, can be anchored onto a carrier or solid phase support. In another embodiment, the reverse situation is possible, in which the probe can be anchored to a solid phase and a sample from a subject can be allowed to react as an unanchored component of the assay.

There are many established methods for anchoring assay components to a solid  
15 phase. These include, without limitation, marker or probe molecules which are immobilized through conjugation of biotin and streptavidin. Such biotinylated assay components can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In  
20 certain embodiments, the surfaces with immobilized assay components can be prepared in advance and stored.

Other suitable carriers or solid phase supports for such assays include any material capable of binding the class of molecule to which the marker or probe belongs. Well-known supports or carriers include, but are not limited to, glass, polystyrene,  
25 nylon, polypropylene, nylon, polyethylene, dextran, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

In order to conduct assays with the above mentioned approaches, the non-immobilized component is added to the solid phase upon which the second component is anchored. After the reaction is complete, uncomplexed components may be removed  
30 (e.g., by washing) under conditions such that any complexes formed will remain immobilized upon the solid phase. The detection of marker/probe complexes anchored to the solid phase can be accomplished in a number of methods outlined herein.

In a preferred embodiment, the probe, when it is the unanchored assay component, can be labeled for the purpose of detection and readout of the assay, either directly or indirectly, with detectable labels discussed herein and which are well-known to one skilled in the art.

5        It is also possible to directly detect marker/probe complex formation without further manipulation or labeling of either component (marker or probe), for example by utilizing the technique of fluorescence energy transfer (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that, upon excitation  
10 with incident light of appropriate wavelength, its emitted fluorescent energy will be absorbed by a fluorescent label on a second 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be  
15 differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, spatial relationships between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through  
20 standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

In another embodiment, determination of the ability of a probe to recognize a marker can be accomplished without labeling either assay component (probe or marker) by utilizing a technology such as real-time Biomolecular Interaction Analysis (BIA) (see, *e.g.*, Sjolander, S. and Urbaniczky, C., 1991, *Anal. Chem.* 63:2338-2345 and  
25 Szabo *et al.*, 1995, *Curr. Opin. Struct. Biol.* 5:699-705). As used herein, "BIA" or "surface plasmon resonance" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)),  
30 resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

Alternatively, in another embodiment, analogous diagnostic and prognostic assays can be conducted with marker and probe as solutes in a liquid phase. In such an assay, the complexed marker and probe are separated from uncomplexed components by any of a number of standard techniques, including but not limited to: differential centrifugation, chromatography, electrophoresis and immunoprecipitation. In differential centrifugation, marker/probe complexes may be separated from uncomplexed assay components through a series of centrifugal steps, due to the different sedimentation equilibria of complexes based on their different sizes and densities (see, for example, Rivas, G., and Minton, A.P., 1993, *Trends Biochem Sci.* 18(8):284-7).

Standard chromatographic techniques may also be utilized to separate complexed molecules from uncomplexed ones. For example, gel filtration chromatography separates molecules based on size, and through the utilization of an appropriate gel filtration resin in a column format, for example, the relatively larger complex may be separated from the relatively smaller uncomplexed components. Similarly, the relatively different charge properties of the marker/probe complex as compared to the uncomplexed components may be exploited to differentiate the complex from uncomplexed components, for example through the utilization of ion-exchange chromatography resins. Such resins and chromatographic techniques are well known to one skilled in the art (see, e.g., Heegaard, N.H., 1998, *J. Mol. Recognit.* Winter 11(1-6):141-8; Hage, D.S., and Tweed, S.A. *J Chromatogr B Biomed Sci Appl* 1997 Oct 10;699(1-2):499-525). Gel electrophoresis may also be employed to separate complexed assay components from unbound components (see, e.g., Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1987-1999). In this technique, protein or nucleic acid complexes are separated based on size or charge, for example. In order to maintain the binding interaction during the electrophoretic process, non-denaturing gel matrix materials and conditions in the absence of reducing agent are typically preferred. Appropriate conditions to the particular assay and components thereof will be well known to one skilled in the art.

In a particular embodiment, the level of mRNA corresponding to the marker can be determined both by *in situ* and by *in vitro* formats in a biological sample using methods known in the art. The term "biological sample" is intended to include tissues, cells, biological fluids and isolates thereof, isolated from a subject, as well as tissues,

cells and fluids present within a subject. Many expression detection methods use isolated RNA. For *in vitro* methods, any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from breast cells (see, e.g., Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York 1987-1999). Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (1989, U.S. Patent No. 4,843,155).

The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a marker of the present invention. Other suitable probes for use in the diagnostic assays of the invention are described herein. Hybridization of an mRNA with the probe indicates that the marker in question is being expressed.

In one format, the mRNA is immobilized on a solid surface and contacted with a probe, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probe(s) are immobilized on a solid surface and the mRNA is contacted with the probe(s), for example, in an Affymetrix gene chip array. A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the markers of the present invention.

An alternative method for determining the level of mRNA corresponding to a marker of the present invention in a sample involves the process of nucleic acid amplification, e.g., by rtPCR (the experimental embodiment set forth in Mullis, 1987, U.S. Patent No. 4,683,202), ligase chain reaction (Barany, 1991, *Proc. Natl. Acad. Sci. USA*, 88:189-193), self sustained sequence replication (Guatelli *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, 1989,

*Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These  
5 detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length  
10 and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

For *in situ* methods, mRNA does not need to be isolated from the breast cells prior to detection. In such methods, a cell or tissue sample is prepared/processed using  
15 known histological methods. The sample is then immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the marker.

As an alternative to making determinations based on the absolute expression level of the marker, determinations may be based on the normalized expression level of the  
20 marker. Expression levels are normalized by correcting the absolute expression level of a marker by comparing its expression to the expression of a gene that is not a marker, e.g., a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene, or epithelial cell-specific genes. This normalization allows the comparison of the expression level in one  
25 sample, e.g., a patient sample, to another sample, e.g., a non-breast cancer sample, or between samples from different sources.

Alternatively, the expression level can be provided as a relative expression level. To determine a relative expression level of a marker, the level of expression of the marker is determined for 10 or more samples of normal versus cancer cell isolates,  
30 preferably 50 or more samples, prior to the determination of the expression level for the sample in question. The mean expression level of each of the genes assayed in the larger number of samples is determined and this is used as a baseline expression level

- 86 -

for the marker. The expression level of the marker determined for the test sample (absolute level of expression) is then divided by the mean expression value obtained for that marker. This provides a relative expression level.

Preferably, the samples used in the baseline determination will be from breast  
5 cancer or from non-breast cancer cells of breast tissue. The choice of the cell source is dependent on the use of the relative expression level. Using expression found in normal tissues as a mean expression score aids in validating whether the marker assayed is breast specific (versus normal cells). In addition, as more data is accumulated, the mean expression value can be revised, providing improved relative expression values based on  
10 accumulated data. Expression data from breast cells provides a means for grading the severity of the breast cancer state.

In another embodiment of the present invention, a polypeptide corresponding to a marker is detected. A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide corresponding to a marker of the invention,  
15 preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe  
20 or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

Proteins from breast cells can be isolated using techniques that are well known to  
25 those of skill in the art. The protein isolation methods employed can, for example, be such as those described in Harlow and Lane (Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

A variety of formats can be employed to determine whether a sample contains a  
30 protein that binds to a given antibody. Examples of such formats include, but are not limited to, enzyme immunoassay (EIA), radioimmunoassay (RIA), Western blot analysis and enzyme linked immunoabsorbant assay (ELISA). A skilled artisan can

readily adapt known protein/antibody detection methods for use in determining whether breast cells express a marker of the present invention.

In one format, antibodies, or antibody fragments, can be used in methods such as Western blots or immunofluorescence techniques to detect the expressed proteins. In  
5 such uses, it is generally preferable to immobilize either the antibody or proteins on a solid support. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

10 One skilled in the art will know many other suitable carriers for binding antibody or antigen, and will be able to adapt such support for use with the present invention. For example, protein isolated from breast cells can be run on a polyacrylamide gel electrophoresis and immobilized onto a solid phase support such as nitrocellulose. The support can then be washed with suitable buffers followed by treatment with the  
15 detectably labeled antibody. The solid phase support can then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support can then be detected by conventional means.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid corresponding to a marker of the invention in a biological sample (*e.g.* an  
20 breast-associated body fluid). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing breast cancer. For example, the kit can comprise a labeled compound or agent capable of detecting a polypeptide or an mRNA encoding a polypeptide corresponding to a marker of the invention in a biological sample and means for determining the amount of the polypeptide or mRNA in  
25 the sample (*e.g.*, an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for interpreting the results obtained using the kit.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a polypeptide corresponding to a  
30 marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable label.



- 88 -

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also comprise, *e.g.*, a buffering agent, a preservative, or a protein stabilizing agent. The kit can further comprise components necessary for detecting the detectable label (*e.g.*, an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

#### B. Pharmacogenomics

Agents or modulators which have a stimulatory or inhibitory effect on expression of a marker of the invention can be administered to individuals to treat (prophylactically or therapeutically) breast cancer in the patient. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the level of expression of a marker of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, *e.g.*, Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic

conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the

5 main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT

10 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among

15 different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic

20 response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the level of expression of a marker of the invention in an individual can be

25 determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure

30 and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of expression of a marker of the invention.

- 90 -

This invention also provides a process for preparing a database comprising at least one of the markers set forth in Tables 1-6. For example, the polynucleotide sequences are stored in a digital storage medium such that a data processing system for standardized representation of the genes that identify a breast cancer cell is compiled.

- 5 The data processing system is useful to analyze gene expression between two cells by first selecting a cell suspected of being of a neoplastic phenotype or genotype and then isolating polynucleotides from the cell. The isolated polynucleotides are sequenced. The sequences from the sample are compared with the sequence(s) present in the database using homology search techniques. Greater than 90%, more preferably greater
- 10 than 95% and more preferably, greater than or equal to 97% sequence identity between the test sequence and the polynucleotides of the present invention is a positive indication that the polynucleotide has been isolated from a breast cancer cell as defined above.

- In an alternative embodiment, the polynucleotides of this invention are sequenced and the information regarding sequence and in some embodiments, relative expression,
- 15 is stored in any functionally relevant program, e.g., in Compare Report using the SAGE software (available through Dr. Ken Kinzler at John Hopkins University). The Compare Report provides a tabulation of the polynucleotide sequences and their abundance for the samples normalized to a defined number of polynucleotides per library (say 25,000). This is then imported into MS-ACCESS either directly or via copying the data into an
- 20 Excel spreadsheet first and then from there into MS-ACCESS for additional manipulations. Other programs such as SYBASE or Oracle that permit the comparison of polynucleotide numbers could be used as alternatives to MS-ACCESS. Enhancements to the software can be designed to incorporate these additional functions. These functions consist in standard Boolean, algebraic, and text search operations,
- 25 applied in various combinations to reduce a large input set of polynucleotides to a manageable subset of a polynucleotide of specifically defined interest.

- One skilled in the art may create groups containing one or more project(s) by combining the counts of specific polynucleotides within a group (e.g.,  $\text{GroupNormal} = \text{Normal1} + \text{Normal2}$ ,  $\text{GroupTumor1} + \text{TumorCellLine}$ ). Additional characteristic values
- 30 are also calculated for each tag in the group (e.g., average count, minimum count, maximum count). One skilled in the art may calculate individual tag count ratios between groups, for example the ratio of the average GroupNormal count to the average

GroupTumor count for each polynucleotide. A statistical measure of the significance of observed differences in tag counts between groups may be calculated.

### C. Monitoring Clinical Trials

5 Monitoring the influence of agents (*e.g.*, drug compounds) on the level of expression of a marker of the invention can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent to affect marker expression can be monitored in clinical trials of subjects receiving treatment for breast cancer. In a preferred embodiment, the present invention provides a method for

10 monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of one or more selected markers of the invention in the pre-administration sample; (iii)

15 obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression of the marker(s) in the post-administration samples; (v) comparing the level of expression of the marker(s) in the pre-administration sample with the level of expression of the marker(s) in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example,

20 increased administration of the agent can be desirable to increase expression of the marker(s) to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent can be desirable to decrease expression of the marker(s) to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

25

### D. Surrogate Markers

The markers of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states, and in particular, breast cancer. As used herein, a "surrogate marker" is an objective

30 biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the

disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The markers of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, antibodies may be employed in an immune-based detection system for a protein marker, or marker-specific radiolabeled probes may be used to detect a mRNA marker.

Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The markers of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA or protein for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in marker DNA may correlate with drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

## VII. Experimental Protocol

### A. Subtracted Libraries and Transcript Profiling

Subtracted libraries are generated using a PCR based method that allows the isolation of clones expressed at higher levels in one population of mRNA (tester) compared to another population (driver). Both tester and driver mRNA populations are converted into cDNA by reverse transcription, and then PCR amplified using the SMART PCR kit from Clontech. Tester and driver cDNAs are then hybridized using the PCR-Select cDNA subtraction kit from Clontech. This technique results in both subtraction and normalization, which is an equalization of copy number of low-

- 94 -

abundance and high-abundance sequences. After generation of the subtractive libraries, a group of 96 or more clones from each library is tested to confirm differential expression by reverse Southern hybridization.

## 5        B. Proteomics

Proteins that are secreted by normal and transformed cells in culture are analyzed to identify those proteins that are likely to be secreted by cancerous cells into body fluids. Supernatants are isolated and MWT-CO filters are used to simplify the mixture of proteins. The proteins are then digested with trypsin. The tryptic peptides are loaded  
10        onto a microcapillary HPLC column where they are separated, and eluted directly into an ion trap mass spectrometer, through a custom-made electrospray ionization source. Throughout the gradient, sequence data is acquired through fragmentation of the four most intense ions (peptides) that elute off the column, while dynamically excluding those that have already been fragmented. In this way, approximately 2000 scans worth  
15        of sequence data are obtained, corresponding to approximately 50 to 200 different proteins in the sample. These data are searched against databases using correlation analysis tools, such as MS-Tag, to identify the proteins in the supernatants.

In addition, protein profiling experiments are undertaken to assess whether the proteins associated with the expression of individual markers of the invention are  
20        secreted. Transcriptional profiling experiments are performed on fractions of RNA that are obtained from either (a) endoplasmic reticulum-associated (ER-associated) ribosomes, or (b) free ribosomes. Eukaryotic RNA which is isolated from ER-associated ribosomes tends to encode secreted and membrane bound proteins rather than intracellular proteins. Accordingly, markers of the invention which exhibit significantly  
25        enhanced expression in fractions of RNA from ER-associated ribosomes (in comparison with RNA from free ribosomes) are predicted to be associated with secreted proteins.

## VIII. Summary Of The Data Provided In The Tables

Table 1 shows 4068 novel nucleotide sequences identified through subtracted  
30        library experiments. The sequences of Table 1 were reinterpreted and those sequences are set forth in Tables 3 and 5. These sequences were determined to be novel through various BLAST searches of the available databases.

The library source for SEQ ID NOS: 1-675 was breast cancer cell cultures (ascites and pleural fluid cultures) versus normal (*i.e.*, non-cancerous) human epithelial mammary cell lines (HMEC). SEQ ID NOS: 1-470 are preferred and SEQ ID NOS: 1-315 are most preferred.

- 5       The library source for SEQ ID NOS: 676-1644 was cancer tissue samples (clinical invasive lobular carcinomas (ILC)) versus normal breast tissue samples. SEQ ID NOS: 676-890 and 1056-1363 are preferred and SEQ ID NOS: 676-792 and 1056-1254 are most preferred.

- 10       The library source for SEQ ID NOS: 1645-2941 was cancer tissue samples (clinical invasive ductal carcinomas (IDC)) versus normal breast tissue samples. SEQ ID NOS: 1645-2454 are preferred and SEQ ID NOS: 1645-2124 are most preferred.

The library source for SEQ ID NOS: 2942-4068 was cancer tissue samples (clinical ductal carcinomas in situ (DCIS)) versus normal breast tissue samples. SEQ ID NOS: 2942-3626 are preferred and SEQ ID NOS: 2942-3351 are most preferred.

- 15       Table 2 shows 4843 novel nucleotide sequences identified through subtracted library experiments. The sequences of Table 2 were reinterpreted and those sequences are set forth in Tables 4 and 5. These sequences were determined to be novel through various BLAST searches of the available databases.

- 20       The tester source for SEQ ID NOS: 1-64, 1960-1976 and 3038-3080 was aggressive breast tumor cell lines and the driver source was indolent breast tumor cell lines (detects markers upregulated in more aggressive tumors).

The tester source for SEQ ID NOS: 65-72, 1879, 1977-2004 and 3081-3127 was indolent breast tumor cell lines and the driver source was aggressive breast tumor cell lines (detects markers upregulated in more indolent tumors).

- 25       The tester source for SEQ ID NOS: 73-629, 1880-1894, 2005-2296 and 3128-3471 was poor clinical outcome breast tumors and the driver source was good clinical outcome breast tumors (detects markers upregulated in more aggressive tumors). "Poor clinical outcome" is defined as the patient suffering disease recurrence following surgery within a period of less than five years. "Good clinical outcome" is defined as  
30       the patient remaining disease free for at least five years or more following surgery.



The tester source for SEQ ID NOS: 630-862, 1895-1900, 2297-2385 and 3472-3602 was good clinical outcome breast tumors and the driver source was poor clinical outcome breast tumors (detects markers upregulated in more indolent tumors).

5 The tester source for SEQ ID NOS: 863-1262, 1901-1910, 2386-2567 and 3602-3988 was breast tumor lymph node metastasis and the driver source was indolent (colloid and tubular) breast tumor samples (detects markers upregulated in more aggressive tumors).

10 The tester source for SEQ ID NOS: 1263-1392, 1911-1916, 2568-2735 and 3989-4319 was indolent (colloid and tubular) breast tumor samples and the driver source was breast tumor lymph node metastasis (detects markers upregulated in more indolent tumors).

15 The tester source for SEQ ID NOS: 1393-1638, 1917-1943, 2736-2940 and 4320-4604 was T1N1 breast tumors (tumors 2.0 cm or less in greatest dimension with regional lymph node metastasis) and the driver source was T1N0 breast tumors (tumors 2.0 cm or less in greatest dimension with no regional lymph node metastasis), good clinical outcome (detects markers upregulated in more aggressive tumors).

The tester source for SEQ ID NOS: 1639-1878, 1944-1959, 2941-3037 and 4605-4843 was T1N0 breast tumors with good clinical outcome and the driver source was T1N1 breast tumors (detects markers upregulated in more indolent tumors).

20 Table 6 shows novel nucleotide sequences shown to be associated with breast cancer.

The contents of all references, patents, published patent applications, and database records cited throughout this application are hereby incorporated by reference.

## 25 Other Embodiments

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

Claims

5

1. An isolated nucleic acid molecule selected from the group consisting of:  
a) a nucleic acid molecule comprising a nucleotide sequence which is at least 90% homologous to a nucleotide sequence of Tables 1-6, or a complement thereof;

10

b) a nucleic acid molecule comprising a fragment of a nucleic acid comprising a nucleotide sequence of Tables 1-6, or a complement thereof; and  
c) a nucleic acid molecule comprising a nucleotide sequence of Tables 1-6, or a complement thereof.

15

2. A vector which contains a nucleic acid molecule of claim 1.

3. A host cell which contains a nucleic acid molecule of claim 1.

20

4. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 90% homologous to a nucleic acid comprising a nucleotide sequence of Tables 1-6.

5. An antibody which selectively binds to a polypeptide of claim 4.

25

6. A method for producing a polypeptide comprising culturing the host cell of claim 3 under conditions in which the nucleic acid molecule is expressed.

- 98 -

7. A method for detecting the presence of a polypeptide of claim 4 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

5 b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 4 in the sample.

8. A kit comprising a compound which selectively binds to the polypeptide of claim 4.

10

9. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

15 b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of claim 1 in the sample.

10. The method of claim 9, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

20

11. The method of claim 9, wherein the sample is isolated from breast tissue.

12. The method of claim 9, wherein the sample is a tumor sample.

25

13. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1.

14. A method of assessing whether a patient is afflicted with breast cancer, the method comprising comparing:
- a) the level of expression of a marker in a patient sample, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6, and
  - 5 b) the normal level of expression of the marker in a control non-breast cancer sample,
- wherein a significant difference between the level of expression of the marker in the patient sample and the normal level is an indication that the patient is afflicted with breast cancer.
- 10
15. The method of claim 14, wherein the marker corresponds to a secreted protein.
16. The method of claim 14, wherein the marker corresponds to a transcribed
- 15 polynucleotide or portion thereof, wherein the polynucleotide comprises the marker.
17. The method of claim 14, wherein the sample comprises cells obtained from the patient.
- 20
18. The method of claim 17, wherein the sample is a breast tissue.
19. The method of claim 17, wherein the cells are in a fluid selected from the group consisting of blood fluid, lymph, ascitic fluid, cystic fluid, urine, a breast exudate and a nipple aspirate.
- 25
20. The method of claim 14, wherein the level of expression of the marker in the sample is assessed by detecting the presence in the sample of a protein corresponding to the marker.
- 30
21. The method of claim 15, wherein the presence of the protein is detected using a reagent which specifically binds with the protein.

- 100 -

22. The method of claim 21, wherein the reagent is selected from the group consisting of an antibody, an antibody derivative, and an antibody fragment.

23. The method of claim 14, wherein the level of expression of the marker in  
5 the sample is assessed by detecting the presence in the sample of a transcribed polynucleotide or portion thereof, wherein the transcribed polynucleotide comprises the marker.

24. The method of claim 23, wherein the transcribed polynucleotide is an  
10 mRNA.

25. The method of claim 23, wherein the transcribed polynucleotide is a cDNA.

15 26. The method of claim 23, wherein the step of detecting further comprises amplifying the transcribed polynucleotide.

27. The method of claim 14, wherein the level of expression of the marker in the sample is assessed by detecting the presence in the sample of a transcribed  
20 polynucleotide which anneals with the marker or anneals with a portion of a polynucleotide wherein the polynucleotide comprises the marker, under stringent hybridization conditions.

28. The method of claim 14, wherein the level of expression of the marker in  
25 the sample differs from the normal level of expression of the marker in a patient not afflicted with breast cancer by a factor of at least about 2.

29. The method of claim 14, wherein the level of expression of the marker in the sample differs from the normal level of expression of the marker in a patient not  
30 afflicted with breast cancer by a factor of at least about 5.

- 101 -

30. The method of claim 14, comprising comparing:
- a) the level of expression in the sample of each of a plurality of markers independently selected from the markers listed in Tables 1-6, and
  - b) the normal level of expression of each of the plurality of markers in
- 5 samples of the same type obtained from control humans not afflicted with breast cancer, wherein the level of expression of more than one of the markers is significantly altered, relative to the corresponding normal levels of expression of the markers, is an indication that the patient is afflicted with breast cancer.
- 10 31. The method of claim 30, wherein the level of expression of each of the markers is significantly altered, relative to the corresponding normal levels of expression of the markers, is an indication that the patient is afflicted with breast cancer.
32. The method of claim 30, wherein the plurality comprises at least three of
- 15 the markers.
33. The method of claim 30, wherein the plurality comprises at least five of the markers.
- 20 34. A method for monitoring the progression of breast cancer in a patient, the method comprising:
- a) detecting in a patient sample at a first point in time, the expression of a marker, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6;
  - 25 b) repeating step a) at a subsequent point in time; and
  - c) comparing the level of expression detected in steps a) and b), and therefrom monitoring the progression of breast cancer in the patient.
35. The method of claim 34, wherein the marker corresponds to a secreted
- 30 protein.

- 102 -

36. The method of claim 34, wherein marker corresponds to a transcribed polynucleotide or portion thereof, wherein the polynucleotide comprises the marker.

37. The method of claim 34, wherein the sample comprises cells obtained  
5 from the patient.

38. The method of claim 34, wherein the patient sample is a breast tissue.

39. The method of claim 34, wherein between the first point in time and the  
10 subsequent point in time, the patient has undergone surgery to remove a tumor.

40. A method of assessing the efficacy of a test compound for inhibiting breast cancer in a patient, the method comprising comparing:

a) expression of a marker in a first sample obtained from the patient and  
15 exposed to the test compound, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6, and

b) expression of the marker in a second sample obtained from the patient, wherein the sample is not exposed to the test compound,

wherein a significantly lower level of expression of the marker in the first  
20 sample, relative to the second sample, is an indication that the test compound is efficacious for inhibiting breast cancer in the patient.

41. The method of claim 40, wherein the first and second samples are portions of a single sample obtained from the patient.

25

42. The method of claim 40, wherein the first and second samples are portions of pooled samples obtained from the patient.

- 103 -

43. A method of assessing the efficacy of a therapy for inhibiting breast cancer in a patient, the method comprising comparing:

a) expression of a marker in the first sample obtained from the patient prior to providing at least a portion of the therapy to the patient, wherein the marker is  
5 selected from the group consisting of the markers listed in Tables 1-6, and

b) expression of the marker in a second sample obtained from the patient following provision of the portion of the therapy,

wherein a significantly lower level of expression of the marker in the second sample, relative to the first sample, is an indication that the therapy is efficacious  
10 for inhibiting breast cancer in the patient.

44. A method of selecting a composition for inhibiting breast cancer in a patient, the method comprising:

a) obtaining a sample comprising cancer cells from the patient;

15 b) separately exposing aliquots of the sample in the presence of a plurality of test compositions;

c) comparing expression of a marker in each of the aliquots, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6; and

d) selecting one of the test compositions which induces a lower level of  
20 expression of the marker in the aliquot containing that test composition, relative to other test compositions.

45. A method of inhibiting breast cancer in a patient, the method comprising:

a) obtaining a sample comprising cancer cells from the patient;

25 b) separately maintaining aliquots of the sample in the presence of a plurality of test compositions;

c) comparing expression of a marker in each of the aliquots, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6; and

d) administering to the patient at least one of the test compositions which  
30 induces a lower level of expression of the marker in the aliquot containing that test composition, relative to other test compositions.



46. A kit for assessing whether a patient is afflicted with breast cancer, the kit comprising reagents for assessing expression of a marker selected from the group consisting of the markers listed in Tables 1-6.
- 5        47. A kit for assessing the presence of breast cancer cells, the kit comprising a nucleic acid probe wherein the probe specifically binds with a transcribed polynucleotide corresponding to a marker selected from the group consisting of the markers listed in Tables 1-6.
- 10       48. A kit for assessing the suitability of each of a plurality of compounds for inhibiting breast cancer in a patient, the kit comprising:
- a) the plurality of compounds; and
  - b) a reagent for assessing expression of a marker selected from the group consisting of the markers listed in Tables 1-6.
- 15       49. A method of making an isolated hybridoma which produces an antibody useful for assessing whether a patient is afflicted with breast cancer, the method comprising:
- isolating a protein or protein fragment corresponding to a marker selected
  - 20 from the group consisting of the markers listed in Tables 1-6;
  - immunizing a mammal using the isolated protein or protein fragment;
  - isolating splenocytes from the immunized mammal;
  - fusing the isolated splenocytes with an immortalized cell line to form
  - hybridomas; and
  - 25 screening individual hybridomas for production of an antibody which specifically binds with the protein or protein fragment to isolate the hybridoma.
50. An antibody produced by a hybridoma made by the method of claim 49.

51. A kit for assessing the presence of human breast cancer cells, the kit comprising an antibody, wherein the antibody specifically binds with a protein corresponding to a marker selected from the group consisting of the markers listed in Tables 1-6.

5

52. A method of assessing the breast cell carcinogenic potential of a test compound, the method comprising:

a) maintaining separate aliquots of breast cells in the presence and absence of the test compound; and

10

b) comparing expression of a marker in each of the aliquots, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6,

wherein a significantly enhanced level of expression of the marker in the aliquot maintained in the presence of the test compound, relative to the aliquot maintained in the absence of the test compound, is an indication that the test compound

15

possesses human breast cell carcinogenic potential.

53. A kit for assessing the breast cell carcinogenic potential of a test compound, the kit comprising breast cells and a reagent for assessing expression of a marker, wherein the marker is selected from the group consisting of the markers listed in Tables 1-6.

20

54. A method of treating a patient afflicted with breast cancer, the method comprising providing to cells of the patient an antisense oligonucleotide complementary to a polynucleotide corresponding to a marker selected from the markers listed in Tables 1-6.

25

55. A method of inhibiting breast cancer in a patient at risk for developing breast cancer, the method comprising inhibiting expression of a gene corresponding to a marker selected from the markers listed in Tables 1-6.

30

**Lawrence**

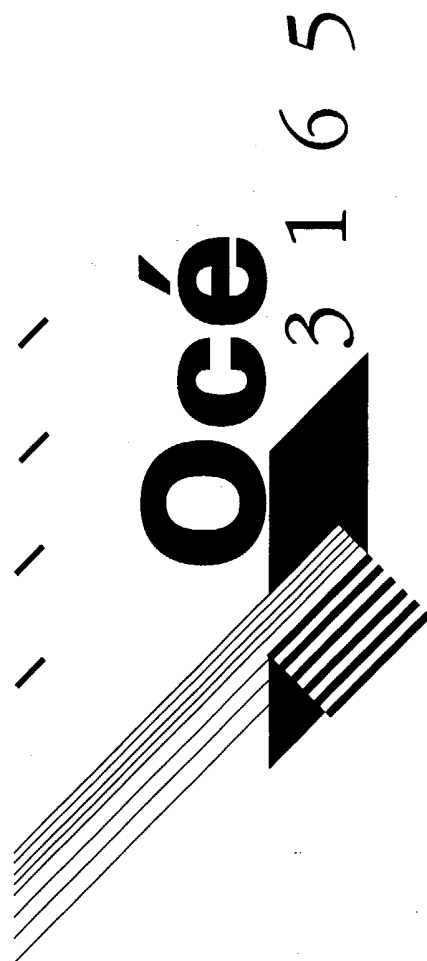
**LARRYT2**

**Wo0151628**

date 02-14-2003

time 14:25:05

printer DAC



Lawrence

LARRYT2

Wo0151628



Table 2

GAACAAGCACAACTGCATGGCCATCAAAAACCCCCAAAAGCATCCCTCTGGTCTGCAAG  
CAGGTTGCTGCACTTCTCCGGTCCTCAATGGGTGCATTTGCAGGCAGCTGATTGCTGTCA  
AATCCAAGGATAGAACTGACATTTTTGTCGTCAAATTCTGGCACTTTGAAACTGTATTCA  
TTAGCTTTAAGGATGCTATTGACTTGTGGAGGTGTGAGGTAAAAATTTCTGTGGTGTGGAA  
GCATATCTCCTTCCTTGGGTGTACCT

Sequence 2013

GTAACAAAGAGACAGCAATATAATAAAAGAGTCTTTCAGTACTGCCTGGGAACAAGCAC  
AACCTGCATGGCCATCAAAAACCCCCAAAAGCATCCCTCTGGTCTGCAAGCAGGTTGCTG  
CACTTCTCCGGTCCTCAATGGGTGCATNNGCAGGCAGCTGATTGCTGTCAAATCCAAGGA  
TAGAACTGACATTTTTGTCGTCAAATTCTGGCACTTTGAAACTGTATTCATTAGCTTTAA  
GGATGCTATTGACTTGTGGAGGTGTGAGGTAAAAATTTCTGTGGTGTGGAAGCATATCTCC  
TTCCTTGGGTGTACCT

Sequence 2014

CCGGGCAGGTACAGATACAGTTTTTAACCTAAAGTTACTGGTGAAACAAGCTTGCTATT  
ACTGAGATGTATAATGAATGTCTTAAAGCTATAAATCCACTCATCGTTGTGTTAAGGAG  
GAACTGAAGCACAAATTGTTAAGAACAAGTGAATTAGGTGAATATACATTGACAAAAGGTT  
GATAATAAAGTTGTCAGAATTTCAAAAATAATTTAGTGTGACTTGAAATCTTAAATCAT  
GTTATATTAATTAAGTAACACTTTACTGATTTAATATTTGAGTCATTTCTAAGGAAAAAT  
ACTGAAATATCAATTGCTTAACAGAAGTTTAAAAATATACGTAATTTGGCATCTTGTTTC  
ACATGTTATGGAAAAGCTAAACATATTTGGGCCTGTTAATTAAAGGCATAAAAAATTATTT  
TATGAGATGGTGTTCATCTGCAAAATACTAACATGATGCACCTTCAAATGCTTACTAATT  
TTCCTAGAAATTAAGGTTACTAAGAGTTAATTAAAAATTAATATTTNGAGTANTTTAAA  
CCTAGAAATANTGAAGGGAACAAATCTGTACCTCGGCCGCTCTAGAACTNNGTGGATCC  
CCCGGGC

Sequence 2015

CCGGGCAGGTACAGATACAGTTTTTAANCCTAAAGTTACTGGTGAAACAAGCTTGCTATT  
ACTGAGATGTATAATGAATGTCTTAAAGCTATAAATCCACTCATCGTTGTGTTAAGGAG  
GAACTGAAGCACAAATTGTTAAGAACAAGTGAATTAGGTGAATATACATTGACAAAAGGTT  
GATAATNAAGTTGTCAGAATTTCAAAAATAATTTAGTGTGACTTGAAATCTTAAATCAT  
GTTATATTAATTAAGTAACACTTTACTGATTTAATATTTGAGTCATTTCTAAGGAAAAAT  
ACTGAAATATCAATTGCTTAACAGAAGTTTAAAAATATACGTAATTTGGCATCTTGTTTC  
ACATGTTATGGAAAAGCTAAACATATTTGGGCCTGTTAATTAAAGGCATAAAAAATTATTT  
TATGAGATGGTGTTCATCTGCAAAATACTAACATGATGCACCTTCAAATGCTTACTAATT  
TTCCTAGAAATTAAGGTTACTAAGAGTTAATTAAAAATTAATATTTAGAGTAATTTAACT  
AGAAATAATTGAAGGGAAACAAA

Sequence 2016

CCGCGGTGGCGGCCCGGGGGCCATTGAGACTACCATGGAAGACTTGAAAGGTCACGTAGCT  
GAGACTTCTGGAGAGACCGTTCAAGGCTTCTGGCTCTTGACAAAGATAGACCACTGGAAC  
AATGAGAAGGAGAGAATTCTACTGGTCACAGACAAGACTCTCTTGATCTGCAAAATACGAC  
TTCATCATGCTGAGTTGTGTGCAGCTGCAGCGGATTCTCTGAGCGCTGTCTATCGCATC  
TGCCTGGGCAAGTTCACCTTCCCTGGGATGTCCCTGGACAAGAGACAAGGAGAAGGCCTT  
AGGATCTACTGGGGGAGTCCGGAGGAGCAGTCTCTTCTTGTCCTGCTGGAACCCATGGTC  
CACTGAAAGTTCTTATGCTACTTTCACTTGAGCATCCTATGAAAATACACCAAGTGAGAA  
ATTCCTTGAAATTTGCAAGTTGTCTGGG

Sequence 2017

ACACCACCTACCTGAATGTCCAGCGGGAAAAATGGGACCATCTCCAGATACCGTGGGAGGC  
CAAGAGCATTTGCTCACTTGCTGATCCTCAGGGACACCAAGACCTACATGCTTGCTTTT  
GACGTGAACGATGAGAAGAACTGGGGGNTGTCTGTCTATGCTGACAAGCCAGAGACGACC  
AAGGAGCAACTGGGAGAGTTCTACGAAGCTCTCGACTGCTTGCGCATTCCCAAGTCAGAT



Table 2

GTCGTGTACCTCGGCCGCCCGGGCAGGTACCACGTCCATGGAGAAGATCTGGACAAGCTC  
CACAGAGCTGCCTGGTGGGGTAAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGAC  
ACTGATGTGAACAAGAGGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAAT  
GGGAATTCAGAAGTAGTAAACTCGTGCTGGACAGACGATGTCAACTTAATGTCCTTGAC  
AACAAAAAGAGGACAGCTCTGACAAAGGCCGTACCTCG

## Sequence 2018

CCGGGCAGGTACTAATCGTTCCTGTTTGATTACTTTTAAACCAGGTGAGAAAAATTAAAT  
TATCGTATTCTAACAAAGTAATATGTGAGATTTTGCAAATGATTTTATAGAAATACACAA  
AATAACTCTTTAGCTTGCTCTGAGCATTTTTCTTTCTGATAGCAACTTTTAAACGTT  
GTGGATCCACAGAACTTACTGCTTTGCTTTCTCTTTGGGGTCATAATCCTCTCCCCTT  
GGAGTGCCACTCCATGCATGTGCACTTAGGATGTGTGGCTGTGTGTGTGTTGGGAACC  
CTCACGGACACATAAGGTTCTATGTCATCAAGTAGAAAACCTATCTCATTATCATTATAA  
TGCTTTCAGATGCTTNTAAGGTTACCTCTTTTTAACATTAGAAGTCAGTGAATGCAG  
CTTTCATTATAATTTTAATACTTTAAATGTTTTGTATTAN

## Sequence 2019

AGGTACTGTTGCAGTGAGCTCAAGTGTTGGGTGTATCAGCTCAAAACACCATGTGATGCC  
AATCATCTCCACAGGAGCAATTTGTTACCAAGAATCTAAGAATTAATCTTAGAATGTA  
TTAATGTTAAATTTCTGTGAGATTATATTGTAGTCACGTAGAATGCCTGACTTGTAGGA  
ATACCCACTTTTTCTGATGCTTTACTAACTTCATCTTTAGATTTAAATCATTAGTA  
GATCCTAGAGGAGCCAGTTTCAGAAAATATAGATTCTAGTTCAGCACCACCCGTAGTTGT  
GCATTGAAATAATTATCATTATGATTATGTATCAGAGCTTCTGGTTTTCTCATTCTTTAT  
TCATTTATTCAACAACCACGTGACAAACACTGGAATTACAGGATGAAGATGAGATAATCC  
GCTCCTTGGCAGTGTTATACTATTATATAACCTGAAAAACAAACAGGTAATTTTCACAC  
AAAGTAATAGATATCATGACACATTTAAATAGGGCACTACTGGAACACACAGATAGGAC  
ATCCAGGTTTTGGGTCAATATTGTAGACTTTTTGGTGGATGAGATATGCAGGTTGATACC  
AGAAGGACAACA

## Sequence 2020

CAGCTCAAAACACCATGTGATGCCAATCATCTCCACAGGAGCAATTTGTTACCAAGAAT  
CTAAGAATTAATCTTAGAATGTATTAATGTTAAATTTCTGTGAGATTATATTGTAGTCA  
CGTAGAATGCCTGACTTGAGGAATACCCANTTTTTCTGATGCTTTACTAACTTCA  
TCTTTTAGATTTAAATCATTAGTAGATCCTAGAGGAGCCAGTTTCAGAAAATATAGATTC  
TAGTTCAGCACCACCCGTAGTTGTGCATTGAAATAATTATCATTATGATTATGTATCAGA  
GCTTCTGGTTTTCTCATTCTTTATTCAATTTATTCAACAACCACGTGACAAACACTGGAAT  
TACAGGATGAAGATGAGATAATCCGCTCCTTGGCAGTGTTATACTATTATATAACCTGAA  
AAAACAAACAGGTAATTTTCACACAAAGTAATAGATATCATGACACATTTAAATAGGGC  
ACTACTGGAACACACAGATAGGACATCCAGGTTTTGGGT

## Sequence 2021

TAGGGCGAATTGGAGCTNACCGCGGCGGCCGCGGTGGCCAGTGAGNCTGCCATGTGCT  
GACTTGAAAGGTCCNTNCTGAGACTTCTGGAGAGACCATTACGGCTTCTGGCTCTTGA  
CAAACATAGACCACTGGAACAATGATTAAGGAGAGAATTCTACTGGCACAGACAAGACTC  
TCTTGATCTGCAAATACGACTTCATCACGCTGAGTTGTGTGCAGCTGCAGCGGATTCCCT  
TGAGCGCTGTCTATCGCATCTGCCTGGGCAAGTTCACCTTCCCTGGGATGTCCCTGGACA  
AGAGACAAGGAGAAGGCCTTAGGATCTACTGGGGGAGTCCGGAGGAGCAGTCTCTTCTGT  
CCCGCTGGAACCCATGGTCCACTGAAGTTCCTTATGCTACTTTCACTGAGCATCCTATGA  
AATACACCAAGTGAGAAATTCCTTGAAATTTGCAAGTTGTCTGGGTTTCATGTCTAAGCTTG  
TTCCAGCTATCCAGAATGCCACAAAGTAATCAACTGGATCTGGAAGAGGAAAGAACTGA  
TGGTGTTAACTGAACC

## Sequence 2022

TGGAANCCACCCCCNTGGCGGCCGCCCGGGCAGGTACNTTTTATCTGAAGGGNAGAGGN





date 02-14-2003

time 14:25:13

printer DAC

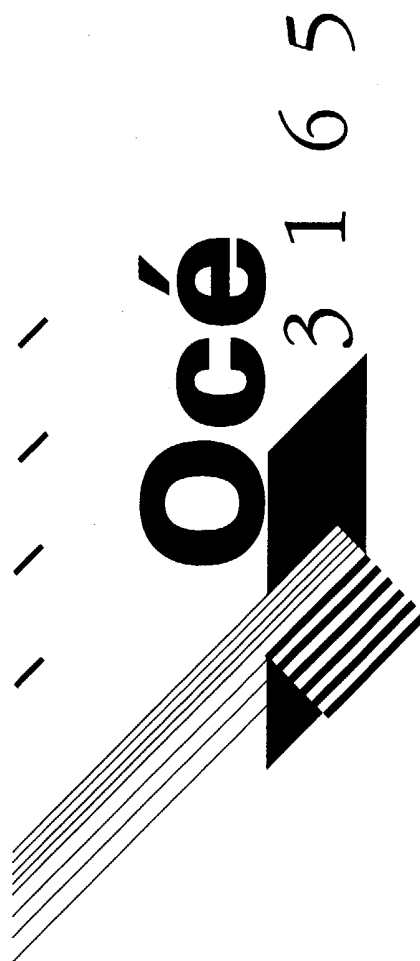




Table 1

TCATTCTGTTACATGACCACAATACAATCAGCACAGATAGATATTAACATTGAGACAATG  
CTCTGATGTAATGGTATCTACTATGCAGTTCATAATCAAATAAGGGTTTTGGTAACTTGG  
GATCCATAAATGGTCTGCAAGGTCAGCCCCGTGTAGACCTGCACTTAGTGGGTGCCA  
Sequence 3381  
AGGTACACCCCTTTTTACCTTGGACAGTTTTTCAGAGAGAAGCCATAGCTGCCTCCTTCC  
TTCACGAGATAGCAGAGCCGGGGCTGGGTCCAAGTTTGCACACCTCCATTATCACAGGG  
GAAAGTATATTATCACTCAAACCTTGCTCCTTCTGACTTTGACCCAACTCTTTCAAGNTC  
CACCCGTGTTTTCACTGCCCTTCTCATAGGAAATCTCCCATCCAGTAACCTAGTTAAAA  
GCTCACNTGGCATTCCCCACTTCTTTCTNGACCAGGATCCACNAACCTGCCATAATGTN  
TCTNTCTTTTGTCCACAAAAGNACACCAATTTGTATCCTTANAGAAACCTCTTGCTTTC  
CATTCTTNGAAAGGCCAGCCCCCTTCTTCTGCNTGGGCCTACACTTCTCAAACCCAACCC  
CGGGACCAAGGGTGGGCCCTCGGGTGT  
Sequence 3382  
AGGTACAATGCCAGGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAA  
ACATTCCAGATGAGTATGGAAATACCACTCTACACTACGCTATCTACAATGAAGATAAAT  
TAATGGCCAAAGCACTGCTCTTATACGGTGCTGATATCGAATCAAAAAACAAGCATGGCC  
TTACCACCACNTGCCACTTGNGTGTACCTGCCCGGGCGGGCCGCTCCTAGNAACCTAGT  
TGGGATCCCCCGGGGGCTGCATGGGAATTCGATATCAANGCTTATCGGATTACCCGTTT  
GACCCNTNGAAGGGGNGGGGNCCCGGTTACCAAGCCTTTTNGNTTCCCTTTTTTAGN  
TNGAAGGGGGTTTAAATNTGCNGCCGCCCTTGGGGCCGTTAAAT  
Sequence 3383  
CCGGGCAGGTACCACCACGCCCTAGATAACTTTTTGTATTTTAAATAGAGACAGAGTTTCA  
CCACGTTGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGATCTGCCTGCCTGGGCCTC  
CCAAAGTGCTGGGATTACAGGCATAAGCCAGCACACCTGGCCGTTAACTAGTTATTTTA  
ATCATGCTACACAAAGAACATTTCTCCAAAGGAAATATATAAAGGGAAACACTGTTGA  
AATAAATGGAAAGCTAAAAGTCTCAGCATAGAAGCTATAAAAGAAACAAGTGGAATG  
TAGAACTGAAAATACAATAACAAAATTTACTGGGAAGGTCTCAACACAATGAAGATGACT  
GAGGGAAGAGTCAGTAAATTTGAAG  
Sequence 3384  
CCGCGGTGGCGGCCGCCCGGGCAGGTACGCGGGGTCACTTACATGAAAATCCCCCTTTGGT  
CACTGAGATCCAAACCTTGCCGGTGTGCTATTGAGCCAGGTCAAGTCTTCGATATGACTG  
TGCTCTACCCTCAACCCAGCCTTTGGCGTTTGTGCTTAGTGACTGCAGAAGCCTCTTA  
GTCTTCCCATCTCCACCTTGCTCTATACCCCTGCCTTGACCCTAGCTGGCCAATAGTAA  
AATCCCCCTCCAAAATATTTTAAAGTAAACAAACAAACAATACAAAATAACGTTTAAAAA  
AAAAAAAAAAAAAAAAAGTCCCTCGGCCGCTCTANAACCTAGGTGGATCCCCCGGGCTGCA  
GGGAATTCG  
Sequence 3385  
CCGCGGTGGCGGCCGCCCGGGCAGGTACTCACAGTTGAGATAGGATTAAACCCCAATTC  
AAAGCCCCCTTACAGAGTTCCACTTACCAGGTTATTTAGCCGGCAACCAGGAAGTAGACT  
GAATGCAATCCTCCCACATCCAAGGTCCCCCGCGTACCT  
Sequence 3386  
CGGCCGCCCGGGCAGGTACACTAAGTGCAGTGCCTCATTTCCTAATTTCCATATAAGCTG  
CCCCTCTGACCTAGAGCATCTTTGTTTTCCACCTACCATAACCTTTAAAGCAAGGGTCCC  
TAACCCCGGGGCCACAGACCCTGTAGGAACCAAGTTGCACAGCAGGAAGTAAGTGGTG  
GGCAAGTGAGCATTCCCGCGTACCT  
Sequence 3387  
CCGGGCAGGTACCGGGCAGCTTGACCTCCATTGCTTTTGGCTTTTGTCTCTTTCTCCTTT  
TGAAGCTCAAAAGGGCATAGAGTGGAAGTCTGATCCTAGGATTTTTTTTCCCTGCTTTGG  
CTGCCTCTGTTTTGGTTCATGTGTCAAGCAGAGACGGGGAAAGCCAAACGACACAATGAG



Table 3

&gt;Sequence 3378

AGGACTTTTTTTTTTTTTTTTTTTTTTTGGATTGTTGATCACACC  
TTTATCTTGCTTAATATGTGTCCAGATTACAAAAATATCTACATCCTTC  
TGATGTATCATTTTAGGGTATGATCTCAGTAATCAGGCAGTGAATTTTCT  
ATAGATTTGTAGGAAGTCTAGGACAAGATGCATATTGTTGGCATGCTGTA  
TACTTCAGAAATGCTTAGTTTATTAATCAACAAAGGAGTCATATCTTTTG  
TCACTAATGTCTCAAAAACAATCCCAACCAATCTTACCGCTAAAAATTGTA  
TTTTATTATTTATTTTTATTTTTACTTTGTAAGCAATCAAAGACAATA  
TGTATCTTATATAGTATAAGGTTTTTTATCTGTAAACTAA

&gt;Sequence 3379

AGGTACGCGGGGGCTTAGACCTGACGCTGGGAGGAGATGCTGCCACCTAG  
GTTACTTGTAGGACCTATACGGCAACCTCCTTTGCCAGGAACATTTTAT  
AAACATCCTGCAGGAAAAAGTCAAGGAAGCTTTTCTTTTGAGCTATTT  
ACAGCTTTTAGCAATTGAGTAAAGTATACTCCTGTGAACAAAAATTGGAA  
CATATTTGTTTCTCTCTAACTGATTTCTCCAGAATTTGGAAGTAGTTAAC  
CAGGAACCCCTCTCTTGGGGTGTGGATCGAGACCCCTTCTGGTAACGAGA  
GCACAGCTTGGTTTTATACATTTTAAGGAGACATGAGACATCAATCAACG  
TATATAAGATGAACATTTTGATCTGAAAAGGCCTTTTTAA

&gt;Sequence 3380

AGGTACTTTTTTTTTTTTTTTTTTTTTTTGCTGGAGAGAGAGAGATGTTT  
TATTTTGGCAAAGCATGTGAGAGTAAGTAGCAAACATCATGACACTTCA  
CCCCTCAAATATTTCCACGTATCACTTAAGAACAAGGGCATTCTTCT  
GTTACATGACCACAATACAATCAGCACAGATAGATATTAACATTGAGACA  
ATGCTCTGATGTATTGGTATCTACTATGCAGTTTATAATCAAATAAGGGT  
TTTGGTAACTTGGGATCCATAAATGGTCTGCAAGGTCAGCCCTGTAGAC  
CTGCACTTAGTGGGTGCCATTTCCCTGTCTCAGGTGCCTGCTCTACTGGT  
GATATTTTCACTGGCTCTGTTGCTCACTTTAAAAATT

&gt;Sequence 3381

AGGTACACCCCTTTTTACCTTGGACAGTTTTTCAGAGAGAAGCCATAGCT  
GCCTCCTTCCTTTCACGAGATAGCAGAGCCGGGGCTGGGTCCAAGTTTGCA  
CACCTCCATTTCATCACAGGGGAAAGTATATTATCACTCAAACCTTGCTCC  
TTCTGACTTTGACCCAACCTTTCAAGTCCACCCGTGTTTTCACTGCTTT  
CTCATAGGAATCCCCATCCAGAACTAGTAAAGTCACTGGATTCCCACTCT  
TTCTGACCAGATCCACAACCTGCATATGTTCTTCTTTGTCCACAAAGACA  
CCATTGATCTCTAAGAAGTCTGTCTCCATCTTGAAGGCCAGCCTTCTCTGC  
TGGGCTACACTTCTCAACCACCCGACCAGGTGGCCCTCGGTGTCCTTCT  
T

&gt;Sequence 3382

AGGTACAATGCCAGGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGC  
ACTGATCCAAACATTCCAGATGAGTATGGAAATACCACTCTACACTACGC  
TATCTACAATGAAGATAAATTAATGGCCAAAGCACTGCTCTTATACGGTG  
CTGATATCGAATCAAAAAACAAGCATGGTCCTCACACCACTGCTACTTGG  
TGTACCTGCCCG

&gt;Sequence 3383

CCGGGCAGGTACCACACGCCTAGATAACTTTTTGTATTTTAAATAGAGA  
CAGAGTTTACCACGTTGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGT  
GATCTGCCTGCTGGGCCCTCCCAAAGTGCTGGGATTACAGGCATAAGCCA  
GCACACCTGGCCGTTAACTAGTTATTTTAATCATGCTACACAAAGAAGA  
CATTTCTCCAAAGGAAATATATAAAGGGAAACACTGTTGAAATAAATGGA  
AAGCTAAAAGTCTCAGCATAGAAGCTATAAAAAGAAACAAGTGGAATTG  
TAGAACTGAAAATACAATAACAAAATTTACTGGAAGGTCTCAACACAATG  
AAGATGACTGAGGGAAGAGTCAGTAAATTTGAAG

&gt;Sequence 3384

CCGGGCAGGTACGCGGGGTCAGTTACATGAAAATCCCCTTTGGTCACTGA  
GATCCAAACCTTGCCGGTGTGTCATTGAGCCAGGTCAAGTCTTCGATATG  
ACTGTGCTCCTACCCTCAACCCAGCCTTTGGCGTTTGTGCCTTAGTGA



Table 1

## Sequence 3919

CCGCGGTGGCGGCCGAGGTACTTTTTTTTTTTAAACAGTTAAGAATGTGTGACGTTATTT  
ATGACCAGAAAGTTCAGATTGAGTCTAAAAATCACAGTAGAGGCTAACCCGTATTATTAT  
TTTTAAACACAATTAACATTTAAAGATGAGTTCGGGGGAGTAATTACTCAGATGTTA  
ACACTAGTGCCCATCAAAGAAGGATTTTAAAAAGATACATTTTCACATTTTCAATTTATGT  
AAAATCTTGATTGTGCTACTCTATTAATACCATGAATAATCTTAAGTAGCAGGTAAGCAA  
ATAAAATACATAATTGTATTTGGCTTATCTGTGTAAGTTACAAATCTGTTTACCTTA  
TTCAACAACAAAATATTTATT

## Sequence 3920

CGAATTGGAGCTCCCCGCGGTGGCGGCCGCCCGGGCAGGTACAATGCCAGGAAGATGAAT  
GTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAACATTCCAGATGAGTATGGAAATA  
CCACTCTACACTACGCTATCTACAATGAAGATAAATTAATGGCCAAAGCACTGCTCTTAT  
ACGGTGCTGATATCGAATCAAAAAACAAGCATGGCCTCACACCACTGCTACTTGGTGTAC  
GCGGGGAAATATCGCCATCGGCATTTCCCCCTACGTTNTTTGGGTGAGTGAATTCAGCC  
TGAGAGAGGAATAGTAAGATACAGTGAAAATGTGAATAGGATCATGTGTTCTTTAATTT  
GAATTATGTATTCTTTTCTTCTATTTTCCAAACTC

## Sequence 3921

GGCGAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACGCGGGTAAACTTTCTGAATAACT  
GAGACATGTCTTAAATTTTGGGAGTTCACAAGGGAACAAGGGTAAGAATTAACCTTCAC  
TAATATTTCTCACCCTCCAAAGCCACTTCAACCACTGCCTTTCTCATTTCCGTTAATAGC  
AACTCCAATCTTCCAGCAGTTTTAGGCTGCAAAACCTTAGCTCCTCTCTGCTTACACTT  
CACATCCAATCTGTACCTATTCAAAGTATCCCCCAAATCGAGCAATTTCTCACCATCTC  
TTACTAATACACCTCAGGTCCAAACCACAGTATACCTCCCTAGAGTACCTGCCCGGGCG  
GCCGCTCTAGAACTAGTTGGATCCCCCGGGCTGCAGGGAATTCGATTATCAAGCTTATCG  
ATACCCGTCGACCTCGAGGGGGGGGCCCGGTACCCANCTTTTTGTTNC

## Sequence 3922

GGAGCTCCCCGCGGTGGCGGCCGAGGTACCAAAGAACAACCGGAAGCCTAAAGATGGACC  
ACAAACCTTTGGATATCCAGGATAACTGAGTTAAAAACCTACCACACGGTTAATAAATA  
ATGATGTCCACTGAGTCCATGACCACAAAGCAAATCACCAGCAGCAGGATGGCATGGGCG  
GCCCTTTTCGCTGGGGGGAGGGACTTGAGCAGAGGCTGGGGCTGTGAAGGTGATGGGAT  
CACCTCTGATGGTTAAACAAAGAGTACCTGCCCGGACCTGCCCG

## Sequence 3923

CCGCCCCGGGCAGGTACANAAAGAGGTTTGTCTGACTTAAAGTGATTTGTCTATCTTTCTTT  
TCCACTCTCCTAATCATGATCTGTTCCCCCTCCTGGCCCAATAACTTAATGCTTGNAAC  
TGAGGTCCACCCATTTATGGAAGNNCTCTTTAGTCCACAGGAGNTCCCTTNAACACCCC  
TTTNTTCTTGTTCCCTGGTNGACAAATTGGTATTCTGGNAANAATTANNCCATNNTTGA  
AAANCCNCCTTTTNNANTAGNGGGTCCGTTCCCNANGTGNTGNACCTANAAATTGNGA  
GTTNAGGGCCAAGATTGAACACCATGTCCANTTTAAAAANGNGGATATANAGNAAANTA  
NNAAAAANATTNNATNNAACAAACNNAGGTANAAATATTCNGTTTNTTTAAAGGNANTC  
NNNATCNTNNTNANNNAGGGCGAACACACGCAATTNCNGTTTTTTTTTNAATTNAAAAA  
AAAAGNTGGNGATTTTNTCTCANAAATTTNACCAACCCAATTAANTGGCCCCCAATTA  
TTNTTNTCAACCTTGGGAAAAATTCCACCCANNNGTGCCAANTACAATTTTNTAAAT  
TTNGGGTAAAAANAATTNGGCCCAACANANCCTTTTTTTAA

## Sequence 3924

CACTACTTAGGGCGAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACTTCAGCTCCACCA  
CGGTGCAGGACAGCTCCTCGTCCCGGAAGTAAGGTATAGTCAGGTGATGCAAACCTCATCA  
TGGGCACCATACCCACTTTTCTTGCTCAAGTAGAACCTTGACTGATGAGTATCTTGGG  
GGTCAAAGGGCATCCCCCATTTGGCTTTAAAGAAGATGTAATTCACCAGGACCATCATTG  
TCTGCGAGTCAAGGTCTTGATCAAGATCTGTGATTTCCCCCTAGTTCCATTCTTCACG





Table 3

ATCTACCGGCTCCTTCTGATGGATTTTGTGTTCTCTTTAGTCAATTCCTT  
CCTGGGGGAGTTTCTGAGGAGAATCATTTGGGATGCAACTGATCACAAGTC  
TTGGCCCTCAGGAGTTTGACATTGCCAGGAACGTTCTAGAAGTATCTAT  
GCACAAACTCTGGTGTGGATTGGCATCTTCTTCTGTCCCCTGCTGCCCTT  
TATCCAAATGATTATGCTTTTCATCATGTTCTACTCCAAAAATATCAGCC  
TGATGATGAATTTCCAGCCTCCGAGCAAAGCCTGGCGGGCCTCACAGATG  
ATGACTTTCTTCATCTTCTTGCTCTTTTCCCATCCTTCACCGGGGTCTT  
GTGCACCTTGCCATCACCATCTTGAGATTGAAGCCTTCAGCTGACTGTG  
GCCCT

>Sequence 3918

CCGGGCAGGTACAGTCCACTAGCACGGCAGCTATAGTGTGGGCATCTAAG  
AGTGCCCTGTAAGCATGTGTGGCCAGGCTAGGGCCATGGGAGAGGCAAGC  
AGACTAAGGAGTGCTGCCATCAGACCAGCCCCATCTCATGTGCAAGACTG  
CCGAGCCTCCAGAGATCAGGTCTCAGAGGAGAAGTCTCTCAAAAGTGAAC  
CCCTGGTCCGCTCAGACACTCTGAAGCCCTAAGGCTGAAATGCCTGGGT  
CGCCCAAACAGCAAAGATGACAATCTGGTCTCCCTGGGAGCTCTGAC  
TCAGTGAGGCCTGAGACCTCTGTGCGCCAGAGAACAGCAGTCAAGGTAGC  
CAGA

>Sequence 3919

AGGTACTTTTTTTTTTAAACAGTTAAGAATGTGTGACGTTATTTATGAC  
CAGAAAGTTCAGATTGAGTCTAAAAATCACAGTAGAGGCTAACCCGTATT  
ATTATTATTTAAACACAATTAACATTTAAAGATGAGTTCTGGGGGAGTA  
ATTACTCAGATGTTAACTAGTGCCCATCAAAGAAGGATTTTAAAAAGA  
TACATTTTCACATTTTATTTTGTAAAACTCTTGATTGTGCTACTCTATT  
AATACCATGAATAATCTTAAGTAGCAGGTAAGCAAATAAAATACATAATT  
GTATTTTGGCTTATTCTGTGTAAGTTACAAATCTGTTTACCTTATTCAA  
CAACAAAATATTATTGG

>Sequence 3920

CCGGGCAGGTACAATGCCAGGAAGATGAATGTGCGTTAATGTTGCTGGAA  
CATGGCACTGATCCAAACATTCCAGATGAGTATGGAAATACCACTCTACA  
CTACGCTATCTACAATGAAGATAAAATTAATGGCCAAAGCACTGCTCTTAT  
ACGGTGTGATATCGAATCAAAAAACAAGCATGGCCTCACACCACTGCTA  
CTTGGTGTACGCGGGGAAATATCGCCATCGGCATTTCCCTACGTTATT  
TGGGAGAGTGAAATTCAGCCTGAGAGAGGAATAGTAAGATACAGTGAAAT  
GTGAATAGATCATGTGTTCTTTTAAATTTGAATTATGTATTCTTTCTCT  
ATTTTCAAACCTGCAATTATCACGCTATTCTTATAACTCTTAAACAAA  
ATAATTGTAGTGAAAAAAGAGGGGTACCT

>Sequence 3921

AGGTACGCGGGTAAACTTTCTGAATTAAGTACGACATGTCTTAAATTTG  
GGAGTTCACAAGGGGAAACAAGGGTAAGAATTAAACTTCACTAATATTTCT  
CACCCTCCAAAGCCACTTCAACCACTGCCTTTCTCATTTTCGGTTAATAGC  
AACTCCAATCTTCCAGCAGGTTTAGGCTGCAAAACCTTAGCTCCTCTCTG  
TCTTACACTTCACATCCAATCTGTCACCTATTCAAAGTATCCCCCAAATC  
GAGCAATTTCTACCATCTCTTACTAATACACCTCAGGTCCAAACCACCA  
GTATACCTCCCTAGAGTACCTGCCCCG

>Sequence 3922

AGGTACCAAAAGAACAAACGGAAGCCTAAAGATGGACCACAAACCTTTGGA  
TATCCAGGATAACTGAGTTAAAAACCTACCACACGGTTAATAAAATAATG  
ATGTCCACTGAGTCCATGACCACAAAGCAAATCACCAGCAGCAGGATGGC  
ATGGGCGGCCCTTTTCGCTGGGGAGGGAAGTGGAGCAGAGGCTGGGGCTGT  
GAAGGTGATGGGATCACCTCTGATGGTTAAACAAGAGTACCTGCCCGGAC  
CTGCCCCG

>Sequence 3923

CCGCCCCGGGCAGGTACAGAAAGAGGTTTGTCTGACTTAAAGTGATTGTG  
ATCTTTCTTTTCCACTCTCCTAATCATGATCTGTTCCCTCTCTGGCCCA  
ATAACTTAATGCTTGTAATACTGAGGTCCACCCATTTATGGAAAACCTT



Table 1

CTTAGGGCGAATTGGAGCTCCCCGCGGTGGCGGCCGCCGGGCAGGTACTCACATCACAA  
TGTCATTTTGATATGGAGCGTTTTGTTATTNTTACAAGGCAGAATGGGGTGNAACAGTTG  
AATTAACTTAGCAATCACGTGCTCANAGCTTTTGCCTGTCAGTTGTGTGTGCCCTTAT  
AGGCCCTTTCCCCACAAGCTCTTTGCTGAAAGAAGTTTGCCTTGNTTTGTTNNGNTGNT  
TTGTATTTANCCANAGGATGTCAAAATTANTCTTCTCAAAGCTTTGAGTAGAGTAAGTGT  
GGGAATAAGCCANTATTTTTTTCTGTTTCTGTAAGCTCAAATGAACGGGTTTTTTCCCT  
TGNNTGCCACTTGNCTAACATGTCCTAAGGTGTTAACCTGCCTCTGACCTGGCTTGC  
AATGCATAGGTGAGGAG

Sequence 3845

AGGGCAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACANCAAAACCCACCTGTGTAAAC  
ACACACAGCAAAGTGATGTAAGAAGTTCCATATAAAGGGCTGCAGTATGGAGAGGTAAT  
GTGCAGGCTGGTTTGCGGCTGTAGGGGCCACCTTGCTGCAGCTCTCCACTGATATGGTAC  
CTGCCCCG

Sequence 3846

GAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACGCGGCCAAAAAGAAGTCAGAAGGAGC  
CAAATCAGGACTGTAAGGTGGATGAATGCCTAATGATTTCCCATCAAAGCTCTCACAAAA  
TTGCCCTTGTTGATGAGAGGAATAAGCAGAANCATTGTGATGGGGAAGGACTCTCTAAT  
GAAGCTTTCCAGGTGTTTTCTGCTAAAGCTTTGGCTTTCTCAAAACACTTTCATAATA  
AGCAGATGTTGTTGTTCTTTGGCCCTCCAGAAAACCAACTAGAAAAATTCCTTTAACATT  
CCAAAAGACTATTGCCATGAGCTTTGTTCTTGACCAGTCCAGTTTGCTTTGACTAGACC  
GCTTCCACTTCTTGGTAGCCATTGCTTTGATTGTGCTTTATCTTTAGGATCATACTGTA  
AAGCCATGTTTTGTCTCCTG

Sequence 3847

ATTGGAGCTCCCCGCGGTGGCAGCGGCCGCCCGGGCAGGTACTGGAATGGCCAAGTGGA  
CTGCTTCAGCTGACCAGGTTCTTTTAAACCGTAGTCATGCTTTCCCACTAACTCTTAAAT  
CCTTATGCTTAGAAAAATTGAGGATAAGGCTGGGCACAGTGGCTCAGGCCTGTAATCCCAG  
CACTTTGGGAGACCAAGGCGGGTGGATCACAAAGGTGAGGAGATCGAGACCATCCTGGCTA  
ACATGGCGAAACCCCGTCTNTACTAAAAATACAAAAAATAGCTGGGCGTGGTGGCGGGC  
GCCTGTANTCCCAGCTACTCATGAGGCTGAGGCAGGAGGATCACTTGAACATGGGAGGCG  
GAGGTTGCAGTGAGCCAAGATGGCGCCACTGCACTCCAGCCTGGGTGATGGCCGTGAGAC  
TCCATGTCAAAAAAAA

Sequence 3848

CTTAGGGCGAATTGGAGCTCCCCGCGGTGGCGGCCGCCCGGGCAAGGTACTTCCGNGCTT  
CTTCATTTCTTCGATAGCCATAAAATTTCTAGCTCTCTATCACCATCCTTATTTATTNC  
TGGTTCTTGAGATCTTTTCTCTGGCTGGCCATTTTCACTGCCTTTAAGCCTTTGTGACTC  
TTCTCTGATGTCAGCTTTAAGNCTNGCTCTGGGTGCTGTNTTCAGAAGAGTATTTTGG  
CATNTGTTTTCTTNGTAGTCAGAACTAATTCGCAAATACTGTGATACTCTTCACTTTC  
GTTGTCAGGAACTGCTGATTTTCAGGTGTTCTGCTCTTCTTTGAGGAATTAATCCATC  
ATNACCATTGCCAGCAGTGACACCATTAGTCAGGTTTTCTAGT

Sequence 3849

GGGCGAANTGGAGCTCCCCGCGGTGGCGGCCGAGGTACTTTTTTTTTTTTTTTTTTTTTT  
TTTGGNCTCCCANGACCTCTTNCACNTTAAAAATGACATCCTAAANGGCTNTTATTTA  
TATGGATATNTACTGATATTGACCATATTCAAATTNAAAAACAAANGNTTGAAAATATTCA  
CTAATTCCTTTAAAAATAGCTCATTNAAAATATGAAAAAGATACTTTTACAAATTNANTG  
AGAATATCACTGNTTNATAATTTGCAAATCNCTTAATGCCTGGCTTANATGAAAACNGC  
TGGAGTCCCATANCNGTTTCTGCATTAAANTTGNTGNTATATGANTTTTNTGCTTAAAC  
CTATGAGGAA

Sequence 3850

CCTATCACTTAGGGCGAATTGGAGCTCCCCGCGGTGGCGGCCGCCCGGGCAGGTACAT



Table 3

## &gt;Sequence 3848

CCGGGCAGGTACTTCCGTGCTTCTTCATTTCTTCGATAGCCATAAAATTT  
 TCTAGCTCTCTATCACCATCCTTATTTATTTCTGGTCTTGAGATCTTTT  
 CTCTGGCTGGCCATTTTCACTGCCTTTAAGCCTTTGTGACTCTTCCTCTG  
 ATGTCAGCTATAAGGCTCGCTCTGGGTTGCTGTTTTCAGAAGAGTATTTT  
 GGCATCTGTTTTCTTTGTAGTCAGAACTAATTCGCAAATCTGTGATA  
 CTCTTCACTTTTCGTTTGTCAAGGAAATTGCTGATTTTCAGGTGTTCTGCTCT  
 TCCTTTGAGGAATTAATCCATCATCACCATTGCCAGCAGTGACACCATT  
 GTCAGGTTTTCTAGTAATCCACATTATTACTTTGATGCTT

## &gt;Sequence 3849

AGGTACTTTTTTTTTTTTTTTTTTTTTTTTGGTCTCCACGACCTCTTT  
 ACACCTTTAAAAATGACATCCTAAACGGCTTTTATTTATATGGATATCTA  
 CTGATATTGACCATATTCAAATTTAAACAAATGTTTGAAATATTTCACT  
 AATTCCTTTAAAAATAGCTCATTAAAAATATGAAAAAGATACTTTTACAA  
 ATTCACTGAGAAATCACTGTTTTATAATTTTGCAAATCTTTAATGCCT  
 GGCTTAAATGAAAACTGCTGGAGTCCCATATCTGTTCTGCATTAAATTT  
 GATGCTATATGATTTTTTTTGCTTAAACATATGAGGAAATGTGGCTTTA  
 CCTAAAAATGTAGTTGTGAAAAGGATAAGGCTTTTAAATAGCCTTTTCTGA  
 TAATTATAGATATTCTG

## &gt;Sequence 3850

CCGGGCAGGTACATCCTGTGAAATCCAGGTGGAGGCTCCCAAACCTCAGT  
 TCTTGACTTCTATGCACCTTGCAGGCTCAACACCACATGAAAGCTGCCAAG  
 GCCTGGGGCTTGAACCTCTGAACGCATGGCCTGAGTTTATACTGGCCCC  
 TTTCAGTCATGGCTGGAGCAGCTGGGATGCAGGGCACCACTTTCTGGAC  
 TGCACACACCATGGGGACCTTGGCCCGGCCACAAAACCATGTTTTCTT  
 CCTAGGCCCTTTGAGCCTGTGATGGGAGGGGCTGCCATGAAGACCTGTGAC  
 ATGTCTGGAGACATTTTCCCATTTGTCTAGGGGATTAAACATTCAGCTCC  
 TCATTACTTTTGCAAATTTCTGCATCCGGCTTGAATTTCTCCTCAGAAAA  
 TGGGATTTTCTTTCTTT

## &gt;Sequence 3851

AGGTACTTGAGGGCTTCATTGCCAGAACACATTATATACAGGATGTCAGA  
 GCTACCAGTGTGCTGCTGGGAGAAAATGCTGCAAAATTCATCTTTTGGAG  
 GGTGGGGGGGAAAACCCAAAAACAACAACAAAAAACTCTCTTACAGAAAT  
 TTCCTTAACATTAAAAAAACTTGTCTATTTTTTCAAAGGCACATTTGAT  
 ACTCAAAATTTGCTAAAAGTATATTTAAAGACATCTAGCCTTCCATATGTA  
 AAAAAATTTTAGATAAAGACAGTTACAGGACTCAGAAATAATATATTGGG  
 TGATCACTGTATCCTCATTTTACATTTAAAGAAATTCAAAATACGTG  
 ATTAATTTTTTTTTTTTAAATGGATGGGTGGTTCCTCGCTTAACCACTTG  
 GGTGCTAGTGGTCTT

## &gt;Sequence 3852

CGAAGGTCGTTGACTCTACTCAGGGCGACTTGGAGCTCCTTCGCGGTGG  
 CGGCCGCCCCGGGCAGGTACGCGGGATCAAAACAAGTCCTCCAAAAGTGGTT  
 GAAATTGGGCATCAAAGGCAAAAACTCCTGAGGGATTGCAAGAATCAGC  
 TAATGTGAAAATCTCAGGAACCTTCAGCAGTGGGTGGACTGTGAGGCTGC  
 CTGAGTTCCAGTTTTCTTCTCAGAACAAAGAATATCCCGATCGGGAAGAT  
 TGCCTACAGAAAAAGGCAAAAAGACCACTGTAGAAACAGAAGATTCTTC  
 TGTAGAGAACCTGAACAGGATCTGTTTGTAGAACAAAAAAAAAAAAAAAA  
 AAAAAAAGTACCT

## &gt;Sequence 3853

GGGGCCATTGAGACTGCCATGGAAGACTTGAAAGGTCACGTATCTGATAC  
 TTCTGGAGAGACCATCAAGGCTTCTGGCTCTTGACAAAGATAGACCACT  
 GGAACAATGAGAAGGAGAGAATTCTACTGGTCACAGACAAGACTCTCTTG  
 ATCTGCAAAATACGACTTCATCATGCTGAGTTGTGTGCAGCTGCAGCGGAT  
 TCCTCTGAGCGTCTCTATCGCATCTGCCTGGGCAAGTTCACCTTCCCTG  
 GGATGTCCCTGGACAAGAGACAAGGAGAAGGCCTTAGGATCTACTGGGGG  
 AGTCCGGAGGAGCAGTCTCTTCTGTCCCGCTGGAACCCATGGTCCACTGA



Table 4

GGATGTCCCTGGACAAGAGACAAGGAGAAGGCCTTAGGATCTACTGGGGG  
AGTCCGGAGGAGCAGTCTCTTCTGTCCCGCTGGAACCCATGGTCCACTGA  
AGTTCCTTATGCTACTTTCACTGAGCATCCTATGAAATACACCAGTGAGA  
AATTCCTTGAAATTTGCAAGTGTCTGGGTTTCATGTCTAAGCTTGTTC  
GCTATCCAGAATGCCACAAGAATCAACTGGATCTGGAAGAGGAAAGAA  
ACTGATGGGGTTAACTGAACCCATTTTGATTGAGACCTGCACAAGGCTGA  
TG

>Sequence 2017

CCGGGCAGGTACCAGACCCGACAGGACCAGTGCATCTATAACACCACCTA  
CCTGAATGTCCAGCGGGAATGGGACCATCTCCAGATCCGTGGGAGGCC  
AAGAGCATTTCGCTCACTTGCTGATCCTCAGGGACACCAAGACCTACATG  
CTTGCTTTTGACGTGAACGATGAGAAGAAGTGGGGGCTGTCTGTCTATGC  
TGACAAGCCAGAGACGACCAAGGAGCAACTGGGAGAGTTCTACGAAGCTC  
TCGACTGCTTGGCGATTCCCAAGTCAGATGTCTGTACCTCGGCCGCCCG  
GGCAGGTACCACGTCCATGGAGAAGATCTGGACAAGCTCCACAGAGCTGC  
CTGGTGGGGTAAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACA  
CTGATGTGAACAAGAGGGACAAGCAAAAGAGGACTGCTCTACATCTGGCC  
TCTGCCAATGGGAATTCAGAAGTAGTAAAC1CGTGCTGGACAGACGATG  
TCAACTTAATGTCTTGACAACAAAAGAGGACAGCTCTGACAAAGGCCG  
TACCT

>Sequence 2018

CCGGGCAGGTACTAATCGTTCCTGTTTGATTACTTTTAAACCAGGTGAGA  
AAAATTAATATGTATTCTAACAAAGTAATATGTGAGATTTTGCAAATG  
ATTTTATAGAAATACACAAAATAACTCTTTAGCTTGCTCTGAGCATT  
TTCTTTTCTGATAGCAACTTTTAAACGTTGTGGATCCACAGAAGTACTG  
CTTTGCTTTCTTTTGGGGTCATAATTCCTCTCCCTTGGAGTGTCCAC  
TCCATGCATGTGCACTTAGGATGTGTGGCTGTGTGTGTGTTGGGAACCC  
TCACGGACACATAAGGTTCTATGTCTCAAGTAGAAAACCTATCTCATT  
TCATTATAATGTCTTCAGATGCTTCTAAGGTTACCTCTTTTTTAAACAT  
TAGAAGTCAGTGAATGCAGCTTTCATTATAATTTTAACTTTAAATG  
TTTTTGATTATGCTGCCATAATGCTCAGCAGCAAAGTTATGACTCACTT  
TAGCAAGTGTGGTAGTTCTTGCTTCAAGCATTGGTTTCATGTAGCTTT  
TATCTTATTTTCTTAGTCTTGTTTTATCAATTACTAAACAATTTG  
TTGATA

>Sequence 2019

AGGTACTGTTGCAGTGAGCTCAAGTGTGGGTGTATCAGCTCAAAACACC  
ATGTGATGCCAATCATCTCCACAGGAGCAATTTGTTTACCAAGAATCTAA  
GAATTAATCTTAGAATGTATTAATGTAAATTTCTGTGAGATTATATTG  
TAGTCACGTAGAATGCTGACTTGTAAGGAATACCCACTTTTTTTCTGA  
TGCTTTACTAATTCATCTTTTAGATTTAAATCATTAGTAGATCCTAGAG  
GAGCCAGTTTCAGAAAATATAGATTCTAGTTCAGCACCACCCGTAGTTGT  
GCATTGAAATAATTATCATTATGATTATGTATCAGAGCTTCTGGTTTCT  
CATCTTTTATTCAATTTATTCAACAACCACGTGACAAACACTGGAATTACA  
GGATGAAGATGAGATAATCCGCTCCTTGGCAGTGTTATACTATTATATAA  
CCTGAAAAAACAACAGGTAATTTTACACAAAGTAATAGATATCATGAC  
ACATTTAAATAGGGCACTACTGGAACACACAGATAGGACATCCAGTTT  
TGGGTCAATATTGTAGACTNTTTGGTGGATGAGATATGCANGTTGATACC  
AGAAGGACAACAN

>Sequence 2020

AGGTACTGTTGCAGTGAGCTCAAGTGTGGGTGTATCAGCTCAAAACACC  
ATGTGATGCCAATCATCTCCACAGGAGCAATTTGTTTACCAAGAATCTAA  
GAATTAATCTTAGAATGTATTAATGTAAATTTCTGTGAGATTATATTG  
TAGTCACGTAGAATGCTGACTTGTAAGGAATACCCACTTTTTTTCTGA  
TGCTTTACTAATTCATCTTTTAGATTTAAATCATTAGTAGATCCTAGAG  
GAGCCAGTTTCAGAAAATATAGATTCTAGTTCAGCACCACCCGTAGTTGT  
GCATTGAAATAATTATCATTATGATTATGTATCAGAGCTTCTGGTTTCT

